AN INTRODUCTION TO THE THEORY OF GENERALIZED MATRIX INVERTIBILITY

by

Thomas L. Boullion

Assistant Professor of Mathematics
University of Southwestern Louisiana
Lafayette, Louisiana

and

Patrick L. Odell

Associate Director
Texas Center for Research
Austin, Texas

This monograph report was prepared with partial support of Computation and Analysis Division of NASA - MSC (Houston) Contract NAS 9-2619.

PREFACE

This is an attempt to assimilate the rapidly increasing literature available on the "pseudo" invertibility of matrices.

Chapter 2 is an exposition of several definitions of a pseudo inverse of a matrix. The equivalence or near equivalence of these definitions are established. Conditions sufficient for the equivalence of others are also given.

Chapter 3 establishes many properties of the Penrose pseudo inverse, which seems to be the formulation most easily understood and lends itself well to algebraic manipulations so that many properties can be established without going into more sophisticated analysis. An attempt has been made to present alternative forms and formulations of the Penrose pseudo inverse while at the same time keeping the presenttation as comprehensible as possible so that a minimum of preparation and effort are required on the part of the reader.

Chapter 4 is devoted to the Scroggs-Odell pseudo inverse which requires more analysis to comprehend and work with than the Penrose definition. Properties of this pseudo inverse are somewhat difficult and lengthy to establish, and thus it is felt that an incorporation of this definition in the previous chapters would disrupt the "minimum of preparation and effort on the part of the reader" attempt in those chapters. Many properties are established and sufficient conditions for others are given. Pseudo inverses in general are investigated and the relationship between any two pseudo inverses is established. Necessary and sufficient conditions for the Scroggs-Odell and Penrose definitions to be equivalent over their common domain of definition are also established.

Chapter 5 is an assimilation of available material on the many applications of pseudo inverses. Emphasis is on applications in the field of statistics and some background in statistics is required.

Chapter 6 is a presentation of several computing schemes for obtaining the (Penrose) pseudoinverse of a matrix. The techniques are presented along with some comment concerning their merits. The same numerical example is used to illustrate several of the techniques, thus facilitating a comparison of the methods presented.

ACKNOWLEDGMENT

We would like to give special acknowledgment with sincere appreciation to the following friends and colleagues for sharing their interest and knowledge with us without which this monograph could never have been written: Professor James Scroggs, University of Arkansas; Professor Truman Lewis, Texas Technological College; Dr. Henry P. Decell, Mr. Al Fievenson and Mr. Mike Speed all of NASA - MSC Houston.

CONTENTS

CHAPTER I

INTRODUCTION

1.1 History

The concept of a generalized inverse for arbitrary $m \times n$ matrices with elements from the real or complex fields is of widespread current interest. Much research has been done in the past decade on the theory and applications of a pseudo inverse for matrices. Indeed, research is presently being carried further at a very rapid pace.

In a paper given at the Fourteenth Western Meeting of the American Mathematical Society at the University of Chicago, April 9-10, 1920, Professor E. H. Moore first called attention to a "useful extension of the classical notion of the reciprocal of a nonsingular square matrix" [65]. In 1935, Moore discussed this concept at some length in his General Analysis [66], his pioneering work being unfortunately somewhat obscured by rather inaccessible notation. Parts of Moore's work have been interpreted by Ben-Israel and Charnes [6] and by Greville [49]. The definition of the pseudo inverse of a matrix A, denoted by A^+, originally given by Moore has been interpreted by Ben-Israel and Charnes [6] to be:

A^+ is the pseudo inverse of A if

$$AA^+ = P_{R(A)} , \tag{1}$$

$$A^+A = P_{R(A^+)} , \tag{2}$$

where $P_{R(A)}$ is an orthogonal projection on the range space of A. Moore established the existence and uniqueness of A^+, for any A, and gave an explicit form for A^+ in terms of the subdeterminants of A and A^*, the conjugate transpose of A. Various properties of A^+ and the relations among A, A^*, and A^+ were incorporated in his General Analysis, and concurrently were given an algebraic basis and extensions by Von Neumann [89] in his studies on regular rings.

Unaware of Moore's results, Bjerhammar [13, 14] and Penrose [70, 71] each gave independent treatments of the pseudo inverse. Bjerhammar constructed A^+ by identifying it with a submatrix of the inverse of a suitable square nonsingular matrix, obtained by multiplying A with another matrix. The general solution of

$$Ax = b ,$$

when solvable, was given by Bjerhammar as

$$x = A^+b + (I - A^+A)y ,$$

where y is arbitrary up to dimensional compatibility. This solution is a corollary of the definition given by Penrose. The least square character of the solution was used by Bjerhammar in geodetic applications; adjusting observations which gave rise to singular matrices.

Penrose [70] defined the pseudo inverse as the unique solution of the equations

$$AXA = A ,$$

$$XAX = X ,$$

$$(AX)^* = AX ,$$

$$(XA)^* = XA .$$

As will be seen in Chapter 2, Penrose's proof of the existence and uniqueness of A^+ is based on the vanishing of a finite polynomial in A^*A.

As mentioned earlier, some of Moore's results did not become well known because the unique notation employed was not adopted by other mathematicians. A clarifying account of Moore's work has been given by Greville [49] where the theory is redeveloped in a clear exposition following the original Moore approach.

A more abstract account of the theory of pseudo inverses has been developed by Ben-Israel and Charnes [6], and F. J. Beutler [11].

Some additional names prominent in the development and study of the theory and applications of pseudo inverses are Hestenes, Tseng, Drazin, Cline, Pyle, Rado, Rao, and Decell.

Many of the researchers in the theory of pseudo inverses have made and discussed various applications of interest and importance. Previously mentioned were the geodetic applications of Bjerhammar. Den Broeder and Charnes, Ben-Israel, and others have given explicit expressions for A^+ as a limit. One of the expressions of den Broeder and Charnes [34],

$$A^+ = \lim_{\lambda \to 0} A^* (\lambda\bar{\lambda} I + AA^*)^{-1} ,$$

was used to solve a problem in diffusion. Other results by den Broeder and Charnes include some theorems on the pseudoinverse, rank, and conditions on nonsingularity for some matrices of special structure, and a necessary and sufficient condition for A to be the solution of the circle composition equation

$$AX = A + X = XA ,$$

where A is normal.

In developing a spectral theory for arbitrary $m \times n$ matrices, which is an extension of Hermitian theory, Hestenes [52] used A^+ in an essential manner to obtain theorems on structure and some properties of matrices relative to "elementary matrices" and the relations of "*-orthogonality" and "*-commutativity."

Penrose suggested applications of the pseudoinverse in least squares solutions to inconsistent linear equations, in particular to statistical problems.

Greville [49] gave an iterative procedure for calculating A^+, using successive partitions of A. Using A^+ he modified the procedure of Dent and Newhouse [35] in constructing polynomials orthogonal over a discrete domain, and used the least squares properties of A^+ in regression analysis.

Pyle [74] and Cline [26], following den Broeder and Charnes, have considered applications to systems of linear equations. The projections AA^+, and A^+A were used by Pyle [74] in a gradient method for solving linear programming problems. These methods were also used by Rosen [78, 79] in his conjugate gradient method for solving linear and nonlinear programs.

An explicit form for the pseudo inverse based upon the Cayley-Hamilton Theorem has been established by Decell [33]. This interesting result leads to a convenient computing technique. The algorithm is outlined briefly in Chapter 6.

Also, Charnes, Cooper, and Thompson [22] have employed the pseudo inverse and the associated solvability criteria in an essential manner to resolve questions of the scope and validity of the so-called "linear programming under uncertainty," and to characterize optimal stochastic decision rules.

Kalman [55] and Florentin [40] have utilized the pseudo inverse in control theory by using its least squares properties in mean square error analysis. Ben-Israel and Charnes [4], following Bott and Duffin [16], have used the pseudo inverse in the analysis of electrical networks, and obtained the explicit solution, dc or ac, in terms of its topological and dynamical characteristics.

1.2 Importance

The role of the pseudo inverse of a matrix is increasing rapidly in importance as the theory of matrices is blossoming in the formulation and solution of problems. Prior to the advent of the electronic computer, a mathematician could talk glibly about the existence and uniqueness of a solution to a system of ten equations in ten unknowns. Few had ever tried to find the solution of such a system. Now matrix theory not only provides an extremely helpful tool for designing a mathematical or statistical model of a system with many variables, but also affords a practical and convenient method of adapting the data for processing by a computer. A problem which occurs in computations resulting in a waste of time and money is trying to compute the inverse of a matrix which is not known in advance to be singular. The concept of a "generalized" or "pseudo" inverse of a matrix overcomes this problem and has been found to be a very useful tool in simplifying and in many cases amplifying the existing theories in many areas of mathematical statistics.

1.3 Reference System

The chapters are divided into numbered sections. Theorems, definitions, etc., are also numbered by chapters. For example, Theorem 2.6 refers to Theorem 6 of Chapter 2.

The equations are numbered anew in each section, and equation numbers are always enclosed in parentheses. Just the equation number is given in referring to an equation in the same section; otherwise chapter and section numbers are prefixed.

Numbers in brackets refer to the numbered references in the Bibliography.

1.4 Basic Concepts and Notation

Capital letters are used to designate matrices and lower case letters for vectors. The n by n identity matrix is denoted by I_n and the null or zero matrix, by $\emptyset$ or simply as 0. Generally, the dimensions are clear from the context. In all cases the dimensions are assumed to be conformable for addition and multiplication to be well-defined. The matrices are assumed to be defined over the field of complex numbers unless specified otherwise. The conjugate transpose A^* of an m by n matrix A is the n by m matrix with ij entry $\bar{a}_{ji}$ where $\bar{a}_{ji}$ is the complex conjugate of the element in the ji position of the matrix A. A matrix is said to be hermitian if $A^* = A$, and normal if $AA^* = A^*A$.

Lower case Greek letters or subscripted lower case letters are used to represent scalars; i.e. Complex or real numbers. If x and y are column vectors, the scalar product $x^*y = (x, y)$ is defined to be $x_1\bar{y}_1 + x_2\bar{y}_2 + \ldots + x_n\bar{y}_n$. If $(x, y) = 0$, the vectors are said to be orthogonal. It is often convenient to denote certain rectangular submatrices of a given matrix by a single letter, and thus to consider matrices whose elements themselves are matrices. A partitioned matrix A in which the submatrices A_{ij} vanish for $i \neq j$ is also called diagonal and is denoted by $A = \text{diag}\ (A_{11}, A_{12}, \ldots, A_{mm})$.

A matrix is said to be invertible or nonsingular if it has an inverse, singular if it does not. A matrix is said to be idempotent if $A^2 = A$. $|A|$ is used to designate the determinant of A. A finite set of matrices is called linearly dependent if there exist scalars, not all zero, such that $\Sigma\alpha_i A_i = 0$. If such a set of scalars does not exist, the set is said to be linearly independent. Linear independence of vectors is a special case of this definition. The rank of a matrix is the maximum number of linearly independent rows or columns of the matrix. For any m by n matrix A having linearly independent columns, there exist n by m matrices B, called left inverses of A, such that $BA = I$. In fact, all such matrices can be characterized in terms of one of them as being expressable in the form $B + U$ where $(B + U)A = I$ and $UA = 0$. A similar situation holds for a matrix with linearly independent rows in terms of a right inverse.

Many times it is advantageous to consider matrices as representations of linear operators on finite dimensional vector spaces. Towards this end, a brief discussion of some linear operator theory needed later on follows. Since every finite dimensional inner-product vector space is a Hilbert space, the setting is assumed to be such a space. For practical purposes one could assume the setting is the Euclidean n-dimensional vector space over the complex number field. The set of all vectors x such that $Ax = 0$ is called the null space of A and is denoted by $N(A)$. The set of all vectors y for which there exist a vector x such that $Ax = y$ is called the range or column space of A, and is denoted by $R(A)$. A vector space X is the direct sum of subspaces U and V if every vector x in X can be written in the form $u + v$, with u in U and v in V, in one and only one way, in which case we write $X = U \oplus V$. If $X = U \oplus V$, the projection on U along V is the transformation E such that $Ex = u$. A linear transformation E is a projection on some subspace if and only if it is idempotent.

Since it is generally clear from the context, no attempt is made to distinguish between a linear operator and its matrix representation.

More specialized notation and definitions are given as needed in the development of the text.

CHAPTER 2

DEFINITIONS AND THEIR RELATIONSHIPS

2.1 The Penrose Definition

Penrose [70] defined the pseudoinverse of any (possibly rectangular) matrix over the field of complex numbers in terms of the unique solution of a certain set of equations. In showing the existence of this matrix, it will be useful to exploit the following properties of the conjugate transpose A^*, of the matrix A.

$$A^{**} = A$$

$$(A+B)^* = A^* + B^*$$

$$(\lambda A)^* = \bar{\lambda} A^*$$

$$(BA)^* = A^* B^*$$

$$AA^* = 0 \quad \text{implies} \quad A = 0 .$$

The last of these follows from the fact that the trace of AA^* is the sum of the moduli of the elements of A. From the last two of these properties we obtain the rule

$$BAA^* = CAA^* \quad \text{implies} \quad BA = CA , \tag{1}$$

since

$$(BAA^* - CAA^*)\,(B - C)^* = (BA - CA)\,(BA - CA)^* .$$

Similarly,

$$BA^*A = CA^*A \quad \text{implies} \quad BA^* = CA^* . \tag{2}$$

Theorem 2.1: The four equations

$$AXA = A \tag{3}$$

$$XAX = X \tag{4}$$

$$(AX)^* = AX \tag{5}$$

$$(XA)^* = XA \tag{6}$$

have a unique solution for any matrix A.

Proof: It will be shown that (4) and (5) are equivalent to the single equation

$$XX^*A^* = X \tag{7}$$

and that (3) and (6) are equivalent to

$$XAA^* = A^* . \tag{8}$$

Equation (7) is obtained by substituting (5) in (4),

$$XAX = X(AX)^* = XX^*A^* = X.$$

Thus, (5) and (4) imply (7). Conversely, (7) implies

$$AXX^*A^* = AX ,$$

the left side of which is hermitian so that

$$AX = (AX)^* .$$

Now by substituting (5) back into (7) we have

$$XX^*A^* = X(AX)^* = XAX = X .$$

Thus, (7) implies (4) and (5).

Substituting (6) into (3) and taking transposes,

$$AXA = A(XA)^* = AA^*X^* = A ,$$

thus

$$(AA^*X^*)^* = XAA^* = A^* .$$

Hence (3) and (6) imply (8). On the other hand, (8) implies

$$XAA^*X^* = A^*X^*$$

in which XAA^*X^* is hermitian so that $(XA)^* = XA$, which is (6).
Now substituting (6) back into (8) gives

$$XAA^* = (XA)^*A^* = A^*X^*A^* = A^* ,$$

and taking transposes again,

$$(A^*X^*A^*)^* = A^*$$

or

$$AXA = A .$$

Therefore (3) and (6) follow from (8).

Summarizing these results:

$XAX = X$ and $(AX)^* = AX$ if and only if $XX^*A^* = X$, and $AXA = A$ and $(XA)^* = XA$ if and only if $XAA^* = A^*$.

It is sufficient then to find a solution X satisfying (7) and (8). Such a matrix will exist if a matrix B can be found such that

$$BA^*AA^* = A^*$$

since $X = BA^*$ will satisfy (8), and since (8) implies

$$A^*X^*A^* = A^*$$

from which it follows that

$$BA^*X^*A^* = BA^*$$

which proves BA^* a solution of (7).

Since A^*A , $(A^*A)^2$, $(A^*A)^3$, . . . cannot all be linearly independent, there exists a relation

$$\lambda_1 A^*A + \lambda_2(A^*A)^2 + \ldots + \lambda_r(A^*A)^r + \lambda_{r+1}(A^*A)^{r+1} +$$
$$+ \ldots + \lambda_k(A^*A)^k = 0 \tag{9}$$

where the λ_i are not all zero. Let λ_r be the first nonzero λ and put

$$B = -\lambda_r^{-1}[\lambda_{r+1}I + \lambda_{r+2}(A^*A) + \ldots + \lambda_k(A^*A)^{k-r-1}] .$$

Then

$$B(A^*A)^{r+1} = -\lambda_r^{-1}[\lambda_{r+1}(A^*A)^{r+1} + \ldots + \lambda_k(A^*A)^k] ,$$

$$B(A^*A)^{r+1} = (A^*A)^r \quad \text{by equation (9).}$$

Now a repeated application of (1) and (2) gives

$$BA^*AA^* = A^* .$$

To show that X is unique, X is assumed to satisfy (7) and (8), which, we recall, summarize the defining equations 3, 4, 5, and 6. Next suppose that Y satisfies equations 3, 4, 5, and 6.

$$AYA = A \qquad (3')$$

$$YAY = Y \qquad (4')$$

$$(AY) = AY \qquad (5')$$

$$(YA) = YA \qquad (6')$$

Substituting (6') in (4') and (5') in (3') gives

$$Y = A^*Y^*Y$$

and

$$A^* = A^*AY .$$

Thus,

$$X = XX^*A^* = XX^*A^*AY = XAY = XAA^*Y^*Y = A^*Y^*Y = Y .$$

The unique solution of equations 3, 4, 5, and 6 was called by Penrose the generalized inverse of the matrix A, denoted A^+ . A more expressive term, pseudoinverse, is used generally in this text, although the terms are used interchangeably. The symbol A^+ , however, has become standard. The conciseness of the Penrose definition, as well as its relative historical priority, makes it well suited for use as a criterion in comparing for equivalence some less susccinctly stated definitions. Definitions due to Moore (interpreted by Greville [49]), Zelen [90], and Frame [41] will be shown to be equivalent by showing that the pseudoinverses defined by these writers satisfy the penrose equations.

2.2 The Greville (Moore) Definition

Greville has developed Moore's definition of the pseudo-inverse of a rectangular matrix by considering first an $m \times n$ $(m \geq n)$ matrix B of maximal rank. Since the columns of B are linearly independent, the vector $v = Bu$ vanishes if and only if u is a zero vector. Therefore, $u^T B^T Bu = v^T v > 0$ whenever $u \neq 0$. Thus $B^T B$ is positive definite and therefore nonsingular. The pseudo-inverse of B is then defined as B^+, where

$$B^+ = (B^T B)^{-1} B^T . \tag{1}$$

Note that this reduces to the ordinary inverse when $m = n$. For $m > n$, the pseudoinverse is a left inverse of B, unique in the sense that it is the only left inverse of B with rows in the row space of B^T. Similarly, the pseudoinverse of an $m \times n$ matrix C of rank m is defined by

$$C^+ = C^T (CC^T)^{-1} . \tag{2}$$

This is the only right inverse of C having columns in the column space of C^T.

Consider the general case of a nonzero matrix A whose rank r may be smaller than its smaller dimension m. Let B denote a matrix of r columns whose columns form a basis for the column space of A. Similarly, let C denote an r rowed matrix whose rows form a basis for the row space of A. The pseudoinverses of B and C are given by (1) and (2), respectively.

Before the pseudoinverse of A is defined, note that A has a unique left identity matrix with rows in the row space of A^T. This is seen to be $I_L = BB^+$, for evidently

$$I_L B = B \tag{3}$$

and it follows that

$$I_L A = A \tag{4}$$

since each column of A is a linear combination of the columns of B. On the other hand, if I_L is of the form XB^T and satisfies (4), it preserves every vector in the column space of A, and we have $XB^TB = B$. Thus $X = B(B^TB)^{-1}$, and $I_L = BB^+$. Similarly,

$$I_R = C^+C \tag{5}$$

is the only matrix with columns in the column space of A which satisfies the relation

$$AI_R = A . \tag{6}$$

It is easily seen that I_L and I_R are both symmetric and idempotent as are $I - I_L$ and $I - I_R$.

The pseudoinverse of any matrix A is now defined to be the unique matrix A^+, which has its rows in the row space of A^T and its columns in the column space of A^T and which satisfies

$$AA^+ = I_L , \quad A^+A = I_R . \tag{7}$$

We investigate the existence of such a matrix A^+ by cases. In the case of matrices of maximal rank it is readily seen that matrices of the type B and C given by (1) and (2) above meet the requirements. In the trivial case of the zero matrix A, if I_L is taken to be the square zero matrix, its rows are in the (null) row space of A^T and its columns are in the (null) column space of A^T, so that equations 4, 6, and 7 are satisfied if we take $A^+ = A^T$.

To show the existence of the pseudoinverse of the general non-zero matrix A, we introduce the matrix H of order r, given by

$$H = B^+AC^+ . \tag{8}$$

Thus

$$BHC = BB^+AC^+C = I_LAI_R = A . \tag{9}$$

Since the rank of a product does not exceed the rank of any factor, (9) shows that H is of rank r, and therefore nonsingular. Finally, we take

$$A^+ = C^+H^{-1}B^+ . \tag{10}$$

It is clear from equations 10, 1, and 2 that this matrix has its rows in the row space of B^T and its columns in the column space of

C^T ; in other words in the row space and column space of A^T . Moreover, since by (9) we have

$$BHC = A$$

and by (10)

$$C^+H^+B^+ = A^+$$

then

$$AA^+ = BHCC^+H^{-1}B = BB^+ = I_L ,$$

and

$$A^+A = C^+H^{-1}B^+BHC = C^+C = I_R ,$$

so that (7) is satisfied.

The following proof of the uniqueness of A^+ appears in Moore's memoir. Suppose A_1^+ and A_2^+ are two matrices satisfying (7) and having their rows and columns in the row and column spaces of A^T . Then

$$A_1^+AA_2^+ = I_RA_2^+ .$$

But $I_R = C^+C$, and the columns of A_2^+ are in the column space of A^T , which is also that of C^+ , so that we can find a matrix X such that $A_2^+ = C^+X$. Therefore,

$$A_1^+AA_2^+ = C^+CC^+X = A_2^+ .$$

Similarly,

$$A_1^+AA_2^+ = A_2^+I_L = A_1^+BB^+ = YB^+BB^+ = A_1^+ ,$$

where $A_1^+ = YB^+$. Thus $A_1^+ = A_2^+$.

2.3 Equivalence of the Penrose and Greville Definitions

Equivalence is established by showing that the Greville pseudoinverse satisfies the Penrose equations. This technique implies complete equivalence because of the uniqueness of the Penrose pseudoinverse. The Greville pseudoinverse is easily shown to be a solution of the Penrose equations by recalling that

$$AA^+ = I_L , \quad A^+A = I_R$$

Where I_L is a left identity of A and I_R is a right identity. Hence, in equation (1.3)

$$AA^+A = I_LA = A .$$

To show that A^+ satisfies (1.4) recall that A^+ has its columns in the column space of A^T, which is also that of C^+ ; so that there exists a matrix X such that $A^+ = C^+X$. Then,

$$A^+AA^+ = I_RA^+ = C^+CC^+X = C^+X = A^+$$

which proves (1.4). Since Greville discussed the Moore pseudoinverse in terms of a matrix with real elements, (1.5) and (1.6) are satisfied if $(AA^+)^T = AA^+$ and $(A^+A)^T = A^+A$. Recall from the Greville definition that

$$AA^+ = I_L = BB^+$$

where B was formed from the linearly independent columns of A . Also recall that $B^+ = (B^TB)^{-1}B^T$ so that

$$AA^+ = I_L = BB^+ = B(B^TB)^{-1}B^T .$$

Then

$$\begin{aligned}(AA^+)^T = I_L^T &= [B(B^TB)^{-1}B^T]^T \\ &= B[B(B^TB)^{-1}]^T \\ &= B(B^TB)^{-1}B^T \\ &= BB^+ \\ &= I_L \\ &= AA^+ .\end{aligned}$$

Similarly, $A^+A = I_R = C^T(CC^T)^{-1}$, where C is formed from the r linearly independent rows of A , is symmetric so that

$$(A^+A)^T = A^+A .$$

2.4 The Zelen Definition

In investigating the role of constraints in the theory of least squares, Zelen [90] finds it adequate for his purpose to develop the pseudoinverse for the less general case of symmetric matrices only.

The result is that the Zelen pseudoinverse if a special case of that of Penrose. As will be pointed out later, a single restriction on the conclusion of Zelen's theorem will suffice to make all the properties of the Penrose pseudoinverse hold for the pseudoinverse defined by Zelen.

Theorem 2.2: *If* A *is a* $p \times p$ *symmetric matrix of rank* q, $q \leq p$, *then there will exist matrices* $H(p \times r)$ *and* $K(p \times r)$ *such that*

$$H^T A = \Phi , \quad |H^T K| \neq 0 . \tag{1}$$

Furthermore, there will exist matrices $C_1(p \times p)$, $C_2(p \times r)$, *and* $C_3(r \times r)$ *such that*

$$\begin{bmatrix} A & K \\ K^T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C_1 & C_2 \\ C_2^T & C_3 \end{bmatrix} \tag{2}$$

having the properties

$$\begin{aligned} &\text{(i)} \quad C_1 \text{ is a symmetric matrix} \\ &\text{(ii)} \quad C_1 = C_1 A C_1 , \quad A = A C_1 A \\ &\text{(iii)} \quad A C_1 = I = K(H^T K)^{-1} H^T \\ &\text{(iv)} \quad C_2 = H(K^T H)^{-1} \\ &\text{(v)} \quad K C_3 = \Phi . \end{aligned} \tag{3}$$

<u>Proof</u>: Since A has rank q, there exist r $(r = p - q)$ linearly independent relations among the rows of A. The nullity of A is r. Thus, if H is formed by selecting as its columns any r linearly independent vectors which form a basis for the null space of A, then

$$AH = \Phi$$

but since A is symmetric,

$$H^T A = \Phi .$$

Let K have as its columns any set of r vectors which form a basis for the null space of A. To show that $H^T K$ is nonsingular, let x be any $r \times 1$ vector and assume that $H^T Kx = 0$. Partitioning H^T into its rows h_i we have that

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_r \end{bmatrix} Kx = \begin{bmatrix} h_1 Kx \\ h_2 Kx \\ \vdots \\ h_r Kx \end{bmatrix} = 0 \quad \text{or} \quad h_i Kx = 0, \; i = 1, 2, \ldots, r .$$

Now $AKx = 0$ since each column of K is a basis vector for the null space of A. But $AKx = 0$ implies that Kx is a vector in the null space of A, and Kx orthogonal to each h_i implies that $Kx = 0$. Now partition K into its columns, say, $K = (k_1, \ldots, k_r)$. Then

$$Kx = (k_1, \ldots, k_r) \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_r \end{bmatrix} = k_1 x_1 + \ldots + k_r x_r = 0 .$$

But the vectors k_i are linearly independent so that $x_1 = x_2 = \ldots = x_r = 0$. Hence, $H^T K$ is nonsingular and has nonvanishing determinant since $H^T Kx = 0$ implies $x = 0$ for any vector x.

Since $|H^T K| \neq 0$, the rows of K^T are linearly independent of the rows of A. This implies that for all vectors $u^T(1 \times p)$ and $v^T(1 \times r)$, $u^T A + v^T K^T = 0$ and hence the augmented matrix

$$\begin{bmatrix} A & K \\ K^T & 0 \end{bmatrix}$$

has full rank. Using the relations of a matrix to its inverse results in

$$\begin{bmatrix} A & K \\ K^T & 0 \end{bmatrix} \begin{bmatrix} C_1 & C_2 \\ C_2^T & C_3 \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_r \end{bmatrix} = \begin{bmatrix} C_1 & C_2 \\ C_2^T & 0 \end{bmatrix} \begin{bmatrix} A & K \\ K^T & 0 \end{bmatrix}$$

or

$$\begin{aligned} &\text{(i)} & AC_1 + KC_2^T &= I_p \\ &\text{(ii)} & K^T C_1 &= \Phi \\ &\text{(iii)} & AC_2 + KC_3 &= \Phi \\ &\text{(iv)} & K^T C_2 &= I_r . \end{aligned} \qquad (4)$$

Equation 4.i implies that

$$H^T A C_1 = H^T K C_2^T = H^T ,$$

but since $|H^T K| \neq 0$, then

$$C_2^T = (H^T K)^{-1} H^T$$

and upon taking the transpose of both sides we get

$$C_2 = H(K^T H)^{-1}$$

which proves (3.iv). Furthermore, substituting (3.iv) into (4.i) gives

$$AC_1 + KC_2^T = I_p$$

$$AC_1 + K(H^T K)^{01} HT = I_p$$

or

$$AC_1 = I_p - K(H^T K)^{-1} H^T$$

which is exactly (3.iii). Now (3.iii) implies that

$$C_1 A C_1 = C_1 - C_1 K(H^T K)^{-1} H^T$$

and applying $C_1 K = \Phi$ gives

$$C_1 A C_1 = C_1 ,$$

thus the first part of (3.ii) holds. Similarly,

$$AC_1 = I - H(K^TH)^{-1}H^T$$

implies

$$AC_1A = A - H(K^TH)^{-1}H^TA = A ,$$

since $H^TA = \Phi$, hence (3.ii). Now since

$$C_2^T = (H^TK)^{-1}H^T$$

then

$$C_2^TA = (H^TK)^{-1}H^TA = \Phi ,$$

and by the symmetry of A ,

$$AC_2 = \Phi .$$

By virtue of (4.iii), $KC_3 = \Phi$, hence (3.v) is proved.

2.5 Equivalence of the Penrose and Zelen Definitions

From (4.3ii) the matrix $C_1 = A^+$ satisfies the first two of the defining equations of Penrose. In order to show that C_1 satisfies the last two Penrose equations, the matrix K must be chosen to be H. This is possible since H and K have the same dimensions and each is formed by having its columns to be any basis for the null space of A. Under this requirement, (4.3iii) becomes

$$AC_1 = I - H(H^TH)^{-1}$$

and (4.3iv) becomes

$$C_2 = H(H^TH)^{-1} .$$

Then

$$(AC_1)^T = [I - H(H^TH)^{-1}H^T]^T = I - H(H^TH)^{-1}H^T = AC_1$$

and $C_1A = I - C_2H^T$ so that

$$(C_1A)^T = [I - H(H^TH)^{-1}H^T]^T = I - H(H^TH)^{-1}H^T = C_1A .$$

2.6 The Frame Definition

The definition of the pseudoinverse given by Frame [41] grows out of his discussion of the solution of degenerate linear systems. Following Frame's example, we will explore in some detail his development of the "semi-inverse" of a matrix and then modify it to the ordinary pseudoinverse. The painstaking approach employed by Frame gives some insight into the application of the pseudoinverse to the method of least squares. It will be useful first to consider some definitions and a theorem on the rank echelon factorization of a matrix.

Definition 2.1: The distinguished columns of a matrix A are the r nonzero columns, no one of which is a linear combination of its predecessors.

Definition 2.2: An $m \times n$ matrix of rank $r \leq m$ is a row echelon matrix if its last $m - r$ rows are zero, its distinguished columns are the first r columns of the unit (identity) matrix I_m, in order, and the 1's in these columns are the first non-zero entries in their respective rows. If $m = r$, there are no rows of zeros and the $r \times n$ matrix is a reduced echelon matrix.

Theorem 2.3: Every $m \times n$ matrix A of rank r has the row echelon factorization

$$\underset{m\times n}{A} = \underset{m\times r}{B}\ \underset{r\times n}{C}$$

where the columns of B are the distinguished columns of A, and C is a reduced echelon matrix.

Proof: Let B_i be the i^{th} distinguished column of A and the i^{th} column of B. Then each column A_j of A can be written

$$A_j = \sum_{i=1}^{r} B_i c_{ij}$$

where the constants of combination c_{ij} form the matrix C. Since each column of A that is not a distinguished column is a linear combination of preceding distinguished columns, C has a reduced echelon form.

The matrix L that converts the matrix A to the row echelon matrix LA is the right-to-left product of the elementary factors $L_1, L_2, \ldots, L_k$;

$$L = L_k \ldots L_2 L_1 ,$$

but it is usually unnecessary to write out these products separately. Indeed, by row operating on (A, I) instead of A, we obtain (LA, L) as the reduced echelon matrix so that if $LA = I$, then $L = A^{-1}$ appears as the right-hand block.

The system $Ax = y$ is called <u>degenerate</u> if the $m \times n$ coefficient matrix A of rank r is not both square and invertible. Either many or no solutions exist. If the vector $y = Ax$ is not zero for any vector x, either $y = Ax$ or some left multiple thereof may still be minimized in length by some vector x_o, using least squares, and the set of soltuions x (if any), or "best fit" vectors x will have the form

$$x = x_o + A_o z$$

where $AA_o = \Phi$, and z is arbitrary.

A matrix A_o of rank $n - r$ is called a <u>complete right annihilator</u> of A if $AA_o = \Phi$. It is the zero matrix if $n = r$. Both the particular vector x_o and a complete right annihilator A_o of A can be read from the partitioned echelon matrix (LA, L) computed by row operations on (A, I).

Let

$$(LA, L) = \begin{bmatrix} L_1A & L_1 \\ & \\ L_2A & L_2 \end{bmatrix} = \begin{bmatrix} C & L_1 \\ & \\ \Phi & L_2 \end{bmatrix}$$

where L_2A is the $(m - r) \times n$ null matrix, and where L_1A is an $r \times n$ reduced echelon matrix,

$$C = L_1A = (I, V)P = (I, V) \begin{bmatrix} P_1 \\ r \times r \\ P_2 \end{bmatrix} = P_1 + VP_2 .$$

The $n \times n$ matrix P is a permutation matrix with inverse $P^* = (P_1^*, P_2^*)$. Its upper $r \times n$ submatrix P_1 has as its nonzero columns all the r distinguished columns of C which are the first r columns of I. The $r \times (n - r)$ matrix V is formed from the remaining columns of C. The rows of the lower submatrix P_2 of P are all the rows of the $n \times n$ unit matrix that do not appear in P , arranged so that

$$C = P_1 + VP_2 .$$

If $r > 0$, the $m \times r$ matrix

$$B = AP_1^*$$

consists of the r distinguished columns of A, and A has the rank factorization

$$A = BC = AP_1^*L_1A . \tag{1}$$

The equations

$$I = P_1P_1^* = (P_1 + VP_2)P_1^* = CP_1^* = L_1AP_1^* = L_1B \tag{2}$$

show that L_1 is a left inverse of B and P_1^* is a right inverse of C. The $n \times m$ matrix $P_1^*L_1 = A^S$ will be called the semi-inverse of the matrix A.

Before the semi-inverse is considered further, we digress to point out incidentally that since $L_2A = \Phi$, a solution x_o of $Ax = y$ can exist only if

$$L_2y = L_2Ax = 0 .$$

In any case, a minimizing vector that reduces the length of L_2y is a solution x_o of $L_1(Ax - y) = \Phi$ and is given by

$$x_o = P_1^*L_1y .$$

Any right annihilator of $A = BC$ also annihilates $C = (I, V)P$. Hence it can be written in the form A_oZ where

$$A_o = P^* \begin{bmatrix} -V \\ \\ I \end{bmatrix} = (P_1^*, P_2^*) \begin{bmatrix} -V \\ \\ I \end{bmatrix} = P_2^* - P_1^*V .$$

The solutions of $Ax - y$ or minimizing vectors x for $L(Ax - y)$ are

$$x = P_1^*L_1^*y + (P_2^* - P_1^*V)z$$

with z arbitrary.

Motivated by equation 1, Frame has stated the following definition:

<u>Definition 2.3</u>: <u>A semi-inverse of an</u> $m \times n$ <u>matrix</u> A <u>of rank</u> r <u>is any</u> $n \times m$ <u>matrix</u> A^s <u>of rank</u> r <u>such that</u>

$$AA^sA = A . \tag{3}$$

If $A = \Phi$, $A^s = A^*$. If A is nonsingular, then $A^s = A^{-1}$, since (3) implies $A^sA = I = AA^s$. Note that both A^sA and AA^s are idempotent since

$$(A^sA)^2 = A^s(AA^sA) = A^sA ,$$

and

$$(AA^s)^2 = (AA^sA)A^s = AA^s .$$

From the above definition, it is clear that the pseudoinverse is a semi-inverse since the Penrose pseuodinverse is included in the set of solutions of (3). On the other hand, every semi-inverse of A satisfies

$$A^sAA^s = A^s$$

since from (1) we have

$$A^sAA = P_1^*L_1AP_1^*L_1$$

but since $L_1A = C$, then from (2)

$$A^sAA^s = P_1^*L_1AP_1^*L_1 = P_1^*CP_1^*L_1 = P_1^*L_1 = A^s .$$

Thus the semi-inverse satisfies the first two Penrose equations. We now have only to examine the circumstances under which both idempotents AA^s and A^sA are hermitian. The restriction on the semi-inverse which accomplishes this is best pointed out in view of a result proved by Frame [41, p. 220].

<u>Theorem 2.4</u>: <u>Every semi-inverse</u> A^s <u>of a matrix</u> $A \neq \Phi$ <u>with rank factorization</u> $A = BC$ <u>has the form</u>

$$A^s = A^{N\ddagger M} = C^{N\ddagger}B^{M\ddagger} = N(MAN)^{\ddagger}M$$

<u>where</u> $CC^{N\ddagger} = I$, N <u>is nonsingular</u> $n \times n$, <u>and</u>

$$C^{N\ddagger} = NN^*C^* \, (CNN^*C^*)^{-1} = N(CN)^{\ddagger} ,$$

<u>which is a right inverse of</u> C, <u>and where</u>

$$B^{\ddagger M} = (B^*M^*MB)^{-1}B^*M^*M = (MB)^{\ddagger}M ,$$

<u>a left inverse of</u> B. <u>The idempotents</u> A^sA <u>and</u> AA^s <u>have the form</u>

$$A^{N\ddagger M}A = C^{N\ddagger}C \quad \text{and} \quad AA^{N\ddagger M} = BB^{\ddagger M} .$$

Now if

$$A^sA = C^{N\ddagger}C = NN^*C^* \, (CNN^*C^*)^{-1}C$$

then A^sA is hermitian if N is chosen to be $I_{n\text{x}n}$, for then

$$A^sA = C^*(CC^*)^{-1}C = [C^*(CC^*)^{-1}C]^* = (A^sA)^* .$$

Similarly, $AA^s = BB^{\ddagger M} = B(B^*M^*MB)^{-1}B^*M^*M$ is hermitian if the non-singular matrix M is chosen to be $I_{m\text{x}m}$.

2.7 The Rao Definition:

The definition of the pseudoinverse by C. R. Rao [77], also grows out of his discussion of the solution of degenerate linear systems. His definition of a "generalized inverse" is given in terms of a consistent system of linear equations.

Definition 2.4: A generalized inverse of a matrix A of order m by n is a matrix of order n by m denoted by A^-, such that for any vector y for which $Ax = y$ is consistent, $x = A^-y$ is a solution.

A generalized inverse so defined is not unique, however, for many applications, as will be pointed out in Chapter 5 this is not necessary.

The equivalence of definitions (2.3) and (2.4) is established in the next theorem.

Theorem 2.5: If A^- is a generalized inverse of A by definition (2.4), then $AA^-A = A$, and conversely.

Proof: Choose y as the i^{th} column a_i of A. Then the equation $Ax = a_i$ is obviously consistent and hence $x = A^-a_i$ is a solution. This implies that $AA^-s_i = a_i$ for all i, which implies that $AA^-A = A$. Conversely, if A^- exists such that $AA^-A = A$ and $Ax = y$ is consistent, then $AA^-Ax = Ax$ or $AA^-y = y$. Hence $x = A^-y$ is a solution.

We now establish how the generalized inverse given by definition (2.4) can be used to obtain the Penrose pseudoinverse. This is the conclusion of the next theorem.

Theorem 2.6: A generalized inverse A^- as given in definition (2.4) can be constructed in such a way that $A^- = A^+$, where A^+ is the Penrose pseudoinverse of A.

Proof: Given A of order m by n, there exists nonsingular, orthogonal matrices P and Q of orders m and n, respectively, such that $PAQ = D$ or $A = P^{-1}DQ^{-1}$ where

$$D = \begin{bmatrix} D_s & 0 \\ 0 & 0 \end{bmatrix}$$

and D_s is a diagonal matrix of order s and rank s. Define $A^- = QD^-P$ where

$$D^- = \begin{bmatrix} D_s^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

Then it follows that $AA^-A = P^{-1}DQ^{-1} = A$. Also $A^-AA^- = QD^-P = A^-$. Also, it is computational to confirm that $(AA^-)^* = AA^-$ and $(A^-A)^* = A^-A$.

2.8 The Desoer and Whalen Definition

The following definition is an extension of the pseudoinverse by Moore and Penrose, and is given from a range-null space point of view. This approach is felt to be beneficial in that the definition has a strong motivation, the concepts are illuminated geometrically, the proofs are quite simple, the basis is eliminated, and the extension to bounded linear mappings with closed range between Hilbert spaces is immediate.

Definition 2.5: Let A be a bounded linear operator of a Hilbert space X into a Hilbert space Y such that $R(A)$ is closed. A^+ is said to be the pseudoinverse of A if

(i) $A^+Ax = x$ for all x in $N(A) = R(A^*)$.

(ii) $A^+y = 0$ for all y in $R(A) = N(A^*)$.

(iii) If $y_1 \varepsilon\ R(A)$ and $y_2 \varepsilon\ N(A^*)$ then $A^+(y_1 + y_2) = A^+y_1 + A^+y_2$.

Since every finite dimensional inner product space is a Hilbert space, it will suffice to show that the above definition is equivalent to the Penrose definition over such a space. Since (i) defines A^+ on $R(A)$, and (ii) defines A^+ on $R(A)$, A^+ is uniquely defined on $Y = R(A) \oplus R(A)$. (i) implies that A^+A is the identity map on $R(A^*)$.

From (iii) we get $AA^+(y_1 + y_2) = AA^+y_1 + AA^+y_2 = AA^+y_1$ so that AA^+ is a projection operator on $R(A)$. Also, if $x = x_1 + x_2$ where $x_1 \varepsilon N(A)$ and $x_2 \varepsilon N(A)^{\perp}$ we have $A^+A(x_1 + x_2) = A^+Ax_1 + A^+Ax_2 = A^+Ax_2$. But $A^+Ax_2 = x_2$ by (i). Hence, A^+A is a projection operator on $N(A) = R(A^*)$. To show that these are orthogonal we establish that AA^+ and A^+A are hermitian. Now $(AA^+)^* = A^{+*}A^*$. Let $x = x_1 + x_2$ where $x_1 \varepsilon N(A^*)$ and $x_2 \varepsilon N(A^*)^{\perp}$, then $(AA^+)^*x = A^{+*}A^*x_1 + A^{+*}A^*x_2 = A^{+*}A^*x_2$. But, (i) implies that $A^{+*}A^*x_2 = x_2$. Now, $AA^+(x_1 + x_2) = AA^+x_1 + AA^+x_2 = AA^+x_2 = x_2$. Hence $(AA^+)^* = AA^+$. Similarly, it can be established that $(A^+A)^* = A^+A$. Interpreting the Penrose equation, $AA^+A = A$, implies that $AA^+AA^+ = AA^+$ and thus that AA^+ is idempotent and hence a projection operator on $R(A)$. Likewise, A^+A is a projection operator on $N(A)$. The fact that those operators are hermitian implies that they are orthogonal projections and thus the Penrose equations could be written more compactly as

$$
\begin{aligned}
AA^+ &= P_{R(A)} \\
A^+A &= P_{R(A^+)}
\end{aligned}
\tag{1}
$$

where P_M is an orthogonal projection on M.

Since the equations in (1) are equivalent to those in the Desoer and Whalen definition for finite dimensional Hilbert spaces, and also equivalent to the Penrose equations in that case, it follows that Definition 2.5 is equivalent to Definition (2.1) in that case. It might be pointed out that equations (1) are essentially those given by Moore in defining a pseudoinverse.

2.9 The Chipman Definition

Before giving this definition we define what is meant by complementary matrices.

Definition 2.6: Two matrices X and Y are said to be complementary if the following two conditions hold:

(i) X and Y both have k columns, and rank X + rank $Y = k$;

(ii) The row space of X and Y have only the origin in common.

Then Y is said to be complementary to X, and vice versa. Further, X and Y are said to be polar if condition (ii) is replaced by the stronger condition

$$(ii') \quad XY' = 0 .$$

(the prime indicating transposition).

This states that the row spaces of X and Y are orthogonal.

Condition (ii) can be written formally as follows: $uX + vY = 0$ implies $uX = vY = 0$. Thus no row of Y (or linear combination thereof) can be linearly dependent on the rows of X, and vice versa. Condition (ii) is implied by condition (ii$'$), since $uX + vY = 0$ and $XY' = 0$ imply

$$vYY'v' = (uX + vY)Y'v' = 0$$

and since $vYY'v'$ is a vanishing sum of squares it follows that $vY = 0$, hence $uX = 0$. Every matrix X has a polar matrix Y (this does not exclude the possibility that Y is the empty $0 \times k$ matrix, or any null matrix with k columns, in case X has rank k); a fortiori, every matrix has a complementary matrix.

For definiteness, let X be $n \times k$ of rank p, and let Y be $m \times k$ of rank q, where $p + q = k$. Let the row spaces of X and Y be denoted $\tilde{X} = \{\xi | \xi = aX\}$ and $\tilde{Y} = \{\eta | \eta = cY\}$; they are of dimension p and q respectively. Every such matrix X possesses a complementary matrix Y, for any $m \geqq q$; for let $\tilde{B} = \{b' \mid Xb = 0\}$ be the q-dimensional column null space of X. Then an $m \times k$ matrix Y can be chosen so that its rows, together with those of X, span $\tilde{X} + \tilde{B} = \tilde{X} + \tilde{Y}$, and so that none of its rows are in $\tilde{X}$; then Y is complementary to X. If the rows of Y are in $\tilde{B}$, then $\tilde{B} = \tilde{Y}$ and Y is polar to X.

<u>Lemma 2.1</u>: Let X and Y be complementary matrices. Then there exist matrices A and B such that $XB = 0$ and rank YB = rank Y, and $YA = 0$ and rank XA = rank X. Moreover, $X(X'X + Y'Y)^{-1}Y' = 0$.

<u>Proof</u>: Let X be $n \times k$ of rank p, and let Y be $m \times k$ of rank q, where $p + q = k$. Define $\nu = n - p$ and $\mu = m - q$. Without loss of generality, let the first p rows X_1 of X have rank p; then the last ν rows X_2 of X may be written $X_2 = NX_1$, where N is $\nu \times p$. Similarly, let the first q rows Y_1 of Y

have rank q; then the last μ rows Y_2 of Y may be written $Y_2 = MY_1$, where M is $\mu \times q$. Since X and Y are complementary, the rows of X_1 and Y_1 form a basis for $\tilde{X} + \tilde{Y}$, and we may define

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}^{-1} = (A_1 \quad B_1)$$

where A_1 and B_1 are $k \times p$ and $k \times q$ respectively. Then

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} (A_1 \quad B_1) = \begin{bmatrix} X_1A_1 & X_1B_1 \\ Y_1A_1 & Y_1B_1 \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}$$

Now define the $k \times \nu$ and $k \times \mu$ matrices A_2 and B_2 by

$$A_2 = A_1N' \qquad\qquad B_2 = B_1M'$$

so that we have

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} I_p \\ N \end{bmatrix} X_1 \quad , \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} I_q \\ M \end{bmatrix} Y_1$$

$$A = [A_1 \quad A_2] = A_1 \, [I_p \quad N'] \quad , \quad B = [B_1 \quad B_2] = B_1 \, [I_q \; M']$$

where A and B are respectively $k \times n$ and $k \times m$. From these relations we obtain

$$XA = \begin{bmatrix} I_p \\ N \end{bmatrix} X_1 A_1 [I_p \; N'] = \begin{bmatrix} I_p & N' \\ N & NN' \end{bmatrix}$$

$$XB = \begin{bmatrix} I_p \\ N \end{bmatrix} X_1 B_1 [I_q \; M'] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$YA = \begin{bmatrix} I_q \\ M \end{bmatrix} Y_1 A_1 [I_p \; N'] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$YB = \begin{bmatrix} I_q \\ M \end{bmatrix} Y_1 B_1 [I_q \; M'] = \begin{bmatrix} I_q & M' \\ M & MM' \end{bmatrix}$$

where rank $XA = p =$ rank X, and rank $YB = q =$ rank Y, proving the first part of the lemma.

To prove that $X(X'X + Y'Y)^{-1}Y' = 0$, first we note that the matrix $W = [{X \atop Y}]$ has rank k (since X and Y are complementary), whence $Q = W'W = X'X + Y'Y$ is positive definite and therefore invertible. From the first part of the lemma we have

$$(X'X + Y'Y)B = Y'YB = Y_1' [I \;\; M'] \begin{bmatrix} I & M' \\ M & MM' \end{bmatrix}$$

Premultiplying by $X(X'X + Y'Y)^{-1} = XQ^{-1}$ we obtain

$$0 = XB = XQ^{-1} Y'YB = XQ^{-1} Y_1' [I \quad M'] \begin{bmatrix} I & M' \\ M & MM' \end{bmatrix}$$

implying $XQ^{-1}Y_1' (I + M'M) = 0$. But $I + M'M$ is positive definite, hence nonsingular, so $XQ^{-1}Y_1' = X(X'X + Y'Y)^{-1}Y_1' = 0$, therefore

$$X(X'X + Y'Y)^{-1}Y' = X(X'X + Y'Y)^{-1}Y_1' [I \quad M'] = 0.$$

In many practical cases one will have $m = q$, in which case the proof of the above lemma becomes greatly simplified, since $YB = I_q$.

We now state the definition given by Chipman [24].

Definition 2.7: Let X^+, A^+ be $n \times k$, $k \times n$ matrices satisfying the equations in theorem 2.1, and let U, V be given symmetric positive definite matrices of orders k and n respectively. Define $X = V^{1/2} X^+ U^{-1/2}$ and $A = U^{1/2} A^+ V^{-1/2}$.

Then

$$\begin{aligned} &\text{(i)} & XAX &= X \\ &\text{(ii)} & AXA &= A \\ &\text{(iii)} & (XA)^T &= V^{-1}XAV \\ &\text{(iv)} & (AX)^T &= V^{-1} AXU \end{aligned} \tag{1}$$

If $U = I_k$ and $V = I_n$, this definition is obviously equivalent to that of Penrose. We shall denote the unique matrix A satisfying the above by $A = X^{\#}$ for given U and V.

The following theorem provides an alternative proof of the existence and uniqueness of the pseudoinverse of a matrix.

<u>Theorem 2.7</u>: Let X and Y be complementary matrices, and define the matrices

$$X^{\#} = (X'X + Y'Y)^{-1}X'$$

$$Y^{\#} = (X'X + Y'Y)^{-1}Y' .$$

Then

(i) $X^{\#}$ and $Y^{\#}$ satisfy properties (i), (ii), and (iv) of Theorem 2.1, and property (iv) of (1) with $U = Q^{-1} = (X'X + Y'Y)^{-1}$.

(ii) $X^{\#}$ and $Y^{\#}$ satisfy property (iv) of (1) for any given U, if and only if $XUY' = 0$.

(iii) In order that $X^{\#}$ (resp. $Y^{\#}$) be unique for any given U, it is necessary and sufficient that Y (resp. X) satisfy $XUY' = 0$.

<u>Proof</u>: i. Defining $W = \begin{bmatrix} X \\ Y \end{bmatrix}$, we have

$$W^{\#} = (W'W)^{-1}W' = (X'X + Y'Y)^{-1}[X'Y'] = [X^{\#}Y^{\#}]$$

whence

$$W^{\#}W = X^{\#}X + Y^{\#}Y = I_k . \qquad (2)$$

From lemma 2.1, $XY^{\#} = X(X'X + Y'Y)^{-1}Y' = (YX^{\#})' = 0$, so successively premultiplying (2) by X and Y, and postmultiplying by $X^{\#}$ and $Y^{\#}$, properties (i) and (ii) of Theorem 2.1 are verified for X and Y. Properties (iv) of Theorem 2.1 and (iv) of (1) are immediately verified, the latter with $U = Q^{-1}$.

ii. Now we show that $X^{\#}XU$ and $Y^{\#}YU$ are symmetric if and only if $XUY' = 0$. Since $Q = X'X + Y'Y$ is symmetric, $X^{\#}XU = Q^{-1}X'XU$ is symmetric if and only if $QX^{\#}XUQ = X^{\#}XUQ$ is symmetric. This in turn is equivalent to the condition that $X'XUY'Y = Y'YUX'X$, which is clearly also necessary and sufficient for the symmetry of $Y^{\#}YU$. (In the case $U = I$, this simply states that $X^{\#}X$ and $Y^{\#}Y$ are symmetric if and only if $X'X$ and $Y'Y$ commute.) Now $XUY' = 0$ implies $X'XUY'Y = 0 = Y'YUX'X$, so the condition $XUY' = 0$ is obviously sufficient.

To see that it is also necessary, observe that $X^{\#'}X'X = (XX^{\#})'X = XX^{\#}X = X$ from (iv) and (i) of Theorem 2.1, and similarly $Y'YY^{\#} = Y'(YY^{\#})' = Y'Y^{\#'}Y' = Y'$.

Therefore

$$X'XUY'Y = Y'YUX'X$$

implies

$$XUY' = X^{\#'}X'XUY'YY^{\#} = X^{\#'}Y'YUX'XY^{\#} = 0$$

since $X^{\#'}Y' = X(X'X + Y'Y)^{-1}Y' = XY^{\#} = 0$ from lemma 2.1.

iii. It is clearly sufficient to show that $X^{\#}$ is unique if and only if Y satisfies $XUY' = 0$. Let Y_1 and Y_2 be two matrices both complementary to X, with row spaces $\tilde{Y}_1$ and $\tilde{Y}_2$ respectively. Define $Q_i = X'X + Y_i'Y_i$ and $X_i^{\#} = Q_i^{-1}X'$ for $i = 1, 2$. In order that $X_1^{\#} = X_2^{\#}$ it is necessary and sufficient that $X' = Q_1Q_2^{-1}X' = X'XQ_2^{-1}X' + Y_1'Y_1Q_2^{-1}X'$. But $X'XQ_2^{-1}X' = X'X_2^{\#'}X' = X'$ from property (i) of Theorem 2.1, so this is equivalent to $Y_1'Y_1Q_2^{-1}X' = 0$. Premultiplying this last equation by $Y_1^{\#'}$, and recalling that $Y_1^{\#'}Y_1'Y_1 = Y_1$ from properties (iv) and (i) of Theorem 2.1, we obtain $Y_1Q_2^{-1}X' = Y_1X_2^{\#} = 0$ as a necessary and sufficient condition that $X_1^{\#} = X_2^{\#}$. Now $Y_2X_2^{\#} = 0$ from lemma 2.1, so Y_1 and Y_2 (which have the same rank) must both be orthogonal to $X_2^{\#'}$ (as well as to $X_1^{\#'}$ by a similar argument). Thus uniqueness is equivalent to the condition $\tilde{Y}_1 = \tilde{Y}_2$, which is guaranteed when Y_1 and Y_2 are both polar to X, in which case $\tilde{Y}_1 = \tilde{Y}_2 = \tilde{B}$ (the column null space of X).

It remains to be shown that the condition $\tilde{Y}_1 = \tilde{Y}_2$ is in turn equivalent to the condition that Y_1 and Y_2 both be orthogonal to XU, i. e., that $Y_1UX' = 0$ and $Y_2UX' = 0$. Let $X_2^{\#}$ satisfy (iv) of (1) for some U; then by assertion ii of the theorem, $Y_2UX' = 0$, and we also have

$$Y_1X_2^{\#}XUX' = Y_1UX'X_2^{\#'}X' = Y_1UX'$$

whence $Y_1X_2^{\#} = 0$ implies $Y_1UX' = 0$. This proves the necessity of the condition $YUX' = 0$. For the sufficiency, assume first that $Y_2UX' = 0$, whence $X_2^\#$ satisfies (iv) of (1) by assertion ii of the theorem; then using property (ii) of Theorem 2.1, we obtain

$$Y_1X_2^\# = Y_1X_2^\#XX_2^\# = Y_1UX'X_2^{\#'}U^{-1}X_2^\#$$

whence $Y_1UX' = 0$ implies $Y_1X_2^\# = 0$, which was to be shown.

We may conclude with a number of remarks concerning this theorem.

Remark 1. The special case of greatest interest is that in which $U = 1$. Then the symmetry of X^+X, which is the equivalent to the commutativity of $X'X$ and $Y'Y$, is in turn equivalent to the orthogonality of X and Y. The condition that Y satisfy $XY' = 0$ is just one way to obtain uniqueness; the essential property is that $X^\#$ is unique with respect to a choice of Y as long as the rows of Y are such as to span a given space $\tilde{Y}$ which is complementary to $\tilde{X}$. This is accomplished equally well by the condition $XUY' = 0$, i.e., that Y be orthogonal to XU. The Moore-Penrose pseudoinverse of X can therefore be defined as the matrix $X^+ = (X'X + Y'Y)^{-1}X'$, where Y is any matrix polar to X. It has the special property that $YX' = YX^+ = 0$, whence the column space of X^+ is the same as the row space of X. On the other hand if $YUX' = YX^\# = 0$, then X is orthogonal to YU but $X^{\#'}$ is orthogonal to Y, so the column space of $X^\#$ is tilted away from the row space of X.

Remark 2. If V and W are any symmetric positive definite matrices of orders n and m respectively, and if $XUY' = 0$, then the matrices

$$X^{\#} = (X'V^{-1}X + Y'W^{-1}Y)^{-1}X'V^{-1}$$

$$Y^{\#} = (X'V^{-1}X + Y'W^{-1}Y)^{-1}Y'W^{-1}$$

satisfy (1) with W replacing V in the case of $Y^{\#}$. This follows immediately by applying Theorem 2.7 to the matrices $\dot{X} = V^{-1/2}X$ and $\dot{Y} = W^{-1/2}Y$.

Remark 3. Theorem 2.7 could just as easily have been established in terms of some $n \times q$ matrix W or rank $v = n - p$, such that $[W \;\; X]$ has rank n. Then $P = WW' + XX'$ has full rank, and the matrix $X'(WW' + XX')^{-1}$ is the generalized inverse of X satisfying (i), (ii), and (iii) of Theorem 2.1, and (iii) of (1) with $V = P^{-1}$.

If $W'X = 0$ and $XY' = 0$ then $X'(WW' + XX')^{-1} = X^{+} = (X'X + Y'Y)^{-1}X'$. For the special case $k = p$ and $q = v$, $[W \;\; X]$ is itself invertible.

Remark 4. Since $XX^{\#}$ is idempotent of rank p and $YY^{\#}$ is idempotent of rank q, if X is $n \times k$ of rank $p = n$, then $XX^{\#} = X(X'X + Y'Y)^{-1}X' = I_p$; and if Y is $m \times k$ of rank $q = m$, then $Y(X'X + Y'Y)^{-1}X' = I_q$. These formulas are useful in applications.

Other formulations of the pseudoinverse of a matrix have appeared in the literature. A formulation due to Scroggs and Odell is given special attention in Chapter 4. Other formulations not included in this chapter will possibly be covered in the properties of the Penrose pseudoinverse or where felt to be so closely related to one of those given to merit not being duplicated. From reading this chapter one might see how to modify the formulations of the pseudoinverses given to meet his own needs.

CHAPTER 3

PROPERTIES

3.1 Elementary Properties of A^+

In this section many properties of the Penrose pseudoinverse of a matrix are given. More elegant and shorter proofs may be obtained in some cases by working with the Desoer and Whalen definition of the pseudoinverse which is given from a range - null - space point of view, however, an attempt is made here to keep this section on as elementary a level as possible so that the results will be comprehended with a minimum of preparation and effort.

We now list two properties of the conjugate transpose of a matrix which will be used frequently in establishing properties of A^+.

a) If A and B are matrices such that AB is defined, then $(AB)^* = B^*A^*$

b) If A is a matrix, then $(A^*)^* = A$.

Theorem 3.1: For any matrix A, the matrix correspondence $A \to A^+$ satisfies the following properties:

P1) $(A^+)^+ = A$.

Proof: By Theorem 2.1, for the n by m matrix A^+ there exists a unique $(m$ by $n)$ matrix $(A^+)^+$ that satisfies the following identities:

$$A^+(A^+)^+ A^+ = A^+$$

$$(A^+)^+A^+ (A^+)^+ = (A^+)^+$$

$$[A^+ (A^+)^+]^* = A^+(A^+)^+$$

$$[(A^+)^+A^+]^* = (A^+)^+A^+$$

However, replacing $(A^+)^+$ by A in the above identities, they become the four defining identities given in Theorem 2.1. Since the matrix X in Theorem 2.1 is unique, it follows that $A = (A^+)^+$.

P2) $(A^*)^+ = (A^+)^* \equiv A^{+*} \equiv A^{*+}$.

<u>Proof</u>: By Theorem 2.1 for the matrix A^*, there exists a unique matrix A^* satisfying the following identities:

$$A^*(A^*)^+ A^* = A^*$$

$$(A^*)^+ A^* (A^*)^+ = (A^*)^+$$

$$[A^*(A^*)^+] = A^*(A^*)^+$$

$$[(A^*)^+ A^*]^* = (A^*)^+A^*$$

$$\begin{aligned}
\text{However,}\quad A^* (A^+)^* A^* &= (AA^+A)^* && \text{Property a of } * . \\
&= A^* && \text{Theorem 2.1} \\
\text{also}\quad (A^+)^*A^*(A^+)^* &= (A^+AA^+)^* && \text{Property a of } * . \\
&= (A^+)^* && \text{Theorem 2.1}
\end{aligned}$$

Likewise, the identities $[(A^+)^* A^*]^* = (A^+)^* A^*$ and $[A^* (A^+)^*]^* = A^* (A^+)^*$ can be verified. Hence, due to the uniqueness, it follows that $(A^+)^* = (A^*)^+$.

P3) $A^+ AA^* = A^*$

<u>Proof</u>: $A^+ AA^* = (A^+ A)^* A^*$ Theorem 2.1

$= (AA^+ A)^*$ Property a of * .

$= A^*$ Theorem 2.1

P4) $A^* AA^+ = A^*$

<u>Proof</u>: $A^* AA^+ = A^* (AA^+)^*$ Theorem 2.1

$= (AA^+ A)^*$ Property a of * .

$= A^*$ Theorem 2.1

P5) $AA^+ A^* = A^{+*}$

<u>Proof</u>: $AA^+ A^{+*} = (AA^+)^* A^{+*}$ Theorem 2.1

$= (A^+ AA^+)^*$ Property a of * .

$= A^{+*}$ Theorem 2.1

P6) $A^{+*} A^+ A = A^{+*}$

<u>Proof</u>: $A^{+*} A^+ A = A^{+*} (A^+ A)^*$ Theorem 2.1

$= (A^+ AA^+)^*$ Property a of * .

$= A^{+*}$ Theorem 2.1

P7) $A^{*+}A^{*}A = A$

Proof: $A^{*+}A^{*}A = (AA^{+})^{*}A$ Property a of $*$.

$= AA^{+}A$ Theorem 2.1

$= A$ Theorem 2.1

P8) $AA^{*}A^{*+} = A$

Proof: $AA^{*}A^{*} = A(A^{+}A)^{*}$ Property a of $*$.

$= AA^{+}A$ Theorem 2.1

$= A$ Theorem 2.1

P9) $A^{*}A^{+*}A^{+} = A^{+}$

Proof: $A^{*}A^{+*}A^{+} = (A^{+}A)^{*}A^{+}$ Property a of $*$.

$= A^{+}AA^{+}$ Theorem 2.1

$= A^{+}$ Theorem 2.1

P10) $A^{+}A^{+*}A^{*} = A^{+}$

Proof: $A^{+}A^{+*}A^{*} = A^{+}(AA^{+})^{*}$ Property a of $*$.

$= A^{+}AA^{+}$ Theorem 2.1

$= A^{+}$ Theorem 2.1

P11) $(AA^*)^+ = A^{+*}A^+$ and $(A^*A)^+ = A^+A^{*+}$

Proof: By Theorem 2.1 there exists a unique matrix $(AA^*)^+$ satisfying the following identities:

$$AA^*(AA^*)^+AA^* = AA^*$$

$$(AA^*)^+AA^*(AA^*)^+ = (AA^*)^+$$

$$[AA^*(AA^*)^+]^* = AA^*(AA^*)^+$$

$$[(AA^*)^+AA^*]^* = (AA^*)^+AA^*$$

It is computational to confirm that replacing $(AA^*)^+$ by $A^{+*}A^+$ in the above yields identities. Hence by the uniqueness of $(AA^*)^+$, the first conclusion follows. The second result is established in a similar manner.

P12) $(AA^*)^+(AA^*) = AA^+$

Proof:

$$\begin{aligned} (AA^*)^+AA^* &= A^{+*}A^+AA^* && \text{P11} \\ &= A^{+*}A^* && \text{P3} \\ &= (AA^+)^* && \text{Property a of } {}^* . \\ &= AA^+ \end{aligned}$$

P13) If $\alpha \neq 0$, then $(\alpha A)^+ = \alpha^{-1}A^+$.

Proof: Direct substitution of $\alpha^{-1}A^+$ into the four defining equations for $(\alpha A)^+$ establishes this result due to the uniqueness.

P14) $0^+ = 0^T$

Proof: For any size null matrix 0, 0^T satisfies the defining equations for 0^+ given in Theorem 2.1. Hence by the uniqueness of 0^+, $0^+ = 0^T$.

P15) If $D = (d_{ij})$ is a square diagonal matrix, then $D^+ = (d^+_{ij})$ where $d^+_{ij} = 0$ for $i \neq j$, $d^+_{ij} = 0$ if $d_{ij} = 0$ and $d^+_{ij} = d^{-1}_{ii}$ if $d_{ii} \neq 0$.

Proof: D^+ as given satisfies the four defining equations in Theorem 2.1 and hence is the unique pseudoinverse of D.

P16) If $A = BC$ where the columns of B are linearly independent and the rows of C are linearly independent, then $A^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$. In particular, if $B = C^*$, $A^+ = C^*(CC^*)^{-2}C$ and $A^+A = AA^+$.

Proof: It is computational to confirm that $X = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$ is a solution of the four defining equations for A^+ in Theorem 2.1. Hence, by the uniqueness of A^+, $X = A^+$. The second part follows by direct substitution also.

P17) $A^+ = (A^*A)^{-1}A^*$ if the columns of A are linearly independent.

Proof: Follows immediately from P16.

P18) $A^+ = A^*(AA^*)^{-1}$ if the rows of A are linearly independent.

Proof: Follows immediately from P16.

P19) $A^{+} = A^{-1}$ if A is square and nonsingular.

Proof: By P18, $A^{+} = A^{*}(AA^{*})^{-1} = A^{*}A^{*-1}A^{-1} = A^{-1}$.

P20) *If* A^{+} *commutes with some power of* A *and* λ *is any nonzero eigenvalue of* A *corresponding to the eigenvector* x, *then* λ^{-1} *is an eigenvalue of* A^{+} *corresponding to the eigenvector* x .

Proof: Let A^{+} commute with A^{n} for some integer $n > 0$, and let $\lambda \neq 0$ be an eigenvalue of A corresponding to the eigenvector x so that

$$Ax = \lambda x,$$

$$x = \lambda^{-1}Ax ,$$

and

$$A^{+}A^{n} = A^{n}A^{+} .$$

Then

$$\begin{aligned} A^{+}x &= \lambda^{-1}A^{+}Ax \\ &= \lambda^{-2}A^{+}A^{2}x \\ &\;\;\vdots \\ &= \lambda^{-n}A^{+}A^{n}x , \end{aligned}$$

by repeated substitution of $\lambda^{-1}Ax$ for x .

Thus

$$\begin{aligned} A^{+}x &= \lambda^{-n}A^{n}A^{+}x \\ &= \lambda^{-n-1}A^{n}A^{+}Ax \\ &= \lambda^{-n-1}A^{n-1}AA^{+}Ax \\ &= \lambda^{-n-1}A^{n-1}Ax \\ &= \lambda^{-1}\lambda^{-n}A^{n}x \\ &= \lambda^{-1}x \, . \end{aligned}$$

Note that this result could be slightly strengthened by replacing the hypothesis that A^{+} commutes with some power of A by $A^{+}A^{n+1} = A^{2}$ for some n.

P21) The row space of A^{+} and A^{*} are identical. Also the column space of A^{+} and A^{*} are identical.

Proof: To establish these results we make use of the fact that if A and B are such that AB is defined, then the row space of AB is contained in the row space of B and the column space of AB is contained in the column space of A. It follows that the row space of $A^{+}A^{+*}A^{*}$ is contained in the row space of A^{*}. However, $A^{+}A^{+*}A^{*} = A^{+}$ by P10, thus, the row space of A^{+} is contained in the row space of A^{*}. Similarly, the row space of $A^{*}AA^{+}$ is contained in the row space of A^{+}. But, $A^{*}AA^{+} = A^{*}$ by P4. Therefore, the row spaces of A^{+} and A^{*} are identical. A similar argument using the equations in P9 and P3 establishes that the column space of A^{+} and A^{*} are identical.

P22) A, A^+ and A^* all have the same rank, $r(A)$.

Proof: Using the fact that the rank of a product is at most the rank of any one of the factors, we have that $r(AA^+A) \leq r(A^+)$. But, $AA^+A = A$ so that $r(A) \leq r(A^+)$. Hence, $r(A) = r(A^+)$. Also $r(A) = R(AA^*A^{*+})$ since by P8, $A = AA^*A^{*+}$. But $r(AA^*A^{*+}) \leq r(A^*)$. Hence, $r(A) \leq r(A^*)$. Now $r(A^*) = r(A^*AA^+)$ since $A^* = A^*AA^+$ by P4. But $r(A^*AA^+) \leq r(A)$. Hence, $r(A^*) \leq r(A)$. It follows that $r(A) = r(A^*)$.

P23) Let A and B be any matrices with the product AB defined. Let $B_1 = A^+AB$ and $A_1 = AB_1B_1^+$. Then $(AB)^+ = B_1^+A_1^+$.

Proof: The product AB can be written as

$$AB = AA^+AB = AB_1 = AB_1B_1^+B_1 = A_1B_1 \ .$$

Let $Y = AB = A_1B_1$ and let $X = B_1^+A_1^+$. Then it is only necessary to show that Y and X satisfy the equations in Theorem 2.1. From the definition of A, we have that $A_1B_1B_1^+ = AB_1B_1^+B_1B_1^+ = A_1$. Now $YX = A_1B_1B_1^+A_1^+ = A_1A_1^+$ is hermitian. Also $YXY = A_1B_1B_1^+A_1^+A_1B_1 = A_1A_1^+A_1B_1 = A_1B_1 = Y$ and

$$XYX = B_1^+A_1^+(A_1B_1B_1^+)A_1^+ = B_1^+A_1^+A_1A_1^+ = B_1^+A_1^+ = X \ .$$

In order to show that XY is hermitian, we observe first that using the definitions of A_1 and B_1 that

$$A^+A_1 = A^+AB_1B_1^+ = A^+A\ (A^+AB)B_1^+ = A^+ABB_1^+ = B_1B_1^+\ .$$

Also, since $A_1^+A_1B_1B_1^+ = A_1A_1^+$, with both A_1^+A and $B_1B_1^+$ hermitian, $B_1B_1^+A_1^+A_1 = A_1^+A_1$. Substituting A^+A_1 for $B_1B_1^+$ gives

$$A_1^+A_1 = A^+A_1A_1^+A_1 = A^+A_1 \text{ and so } A_1^+A_1 = B_1B_1^+\ .$$

From this it now follows that $XY = B_1^+A_1^+A_1B = B_1^+B_1B_1^+B_1 = B_1^+B_1$ is hermitian. Since it has been shown that Y and X satisfy the defining equations for the Penrose pseudoinverse, $X = Y^+$. But $X = B_1^+A_1^+$.

P24) If $A^*A = PDP^*$, where $PP^* = P^*P = I$, and D is diagonal, then $A^+ = PD^+P^*A^*$.

<u>Proof</u>: Suppose $A^*A = PDP^*$, then $(A^*A)^+ = (PDP^*)^+$. Letting $X = PD^+P^*$ in the defining equations for $(PDP^*)^+$ we have

$$PDP^*PD^+P^*PDP^* = PDD^+DP^* = PDP^*\ ,$$
$$PD^+P^*PDP^*PD^+P^* = PD^+DD^+P^* = PD^+P^*\ ,$$

$$[PD^+P^*PDP^*]^* = [PD^+DP^*]^* = P(D^+D)^*P^* = PD^+DP^* = PD^+P^*PDP^*\ ,$$

and

$$[PDP^*PD^+P^*]^* = [PDD^+P^*]^* = P(DD^+)^*P^* = PDD^+P^* = PDP^*PD^+P^*\ .$$

Hence $\quad (PDP^*)^+ = PD^+P^*$

and thus

$$(A^*A)^+ = PD^+P^*\ .$$

Multiplying by A^* and noting that $(A^*A)^+ = A^+A^{*+}$ by P11 we have

$$A^+A^{*+}A^* = PD^+P^*A^* .$$

But by P2 and P10 the left member is A^+ . Hence the conclusion that

$$A^+ = PD^+P^*A^* .$$

P25) If $A = \Sigma A_i$, where $A_iA_j^* = 0$ and $A_i^*A_j = 0$ whenever $i \neq j$, then

$$A^+ = \Sigma A_i^+ .$$

Proof: Assume $A_iA_j^* = 0$ and $A_i^*A_j = 0$ whenever $i \neq j$. Then, since $A_j^+ = A_j^*A_j^{+*}A_j^+$ and $A_i^+ = A_i^+A_i^{+*}A_i^*$ it follows that $A_i^+A_j = 0$ and $A_iA_j^+ = 0$. This implies that all the cross product terms in AA^+ are zero, and thus $AA^+ = \Sigma A_iA_i^+$, $A^+A = \Sigma A_i^+A_i$. Hence, by direct computation the four defining equations for A^+ are found to be satisfied by ΣA_i^+ . By the uniqueness of A^+ we have that $A^+ = \Sigma A_i^+$.

P26) If A is normal, $A^+A = AA^+$ and $(A^n)^+ = (A^+)^n$.

Proof: By P10 $A^+A^{+*}A^* = A^+$. Applying P2 to this gives $A^+A^{*+}A^* = A^+$ and P11 gives $(A^*A)^+A^*A = A^+A$. Also, since A is normal $A^*A = AA^*$ so that $A^+A = (AA^*)^+AA^*$. By P11 $(AA^*)^+ = A^{*+}A^+$ and $A^{*+}A^+ = A^{+*}$ by P2. Hence, since $A^{+*}A^+A = A$ by P6, we have $A^+A = A^{*+}A^+AA^* = A^{+*}A^* = (AA^+) = AA^+$. To establish the second part we note that the first part implies that $(A^+A)^n = A^n(A^+)^n = (A^+)^nA^n$. Direct substitution of $(A^+)^n$ in the defining equations for $(A^n)^+$ yields the desired result.

P27) $AB = 0$ if and only if $B^+A^+ = 0$.

Proof: Assume $AB = 0$. Premultiplying by A^+ and post-multiplying by B^+ we get $A^+ABB^+ = 0$. Taking conjugate transposes, $(BB^+)^*(A^+A)^* = BB^+A^+A = 0$. Multiplying by B^+ and A^+ on the left and right, respectively gives $B^+BB^+A^+AA^+ = B^+A^+ = 0$. Conversely, if $B^+A^+ = 0$ we have that $BB^+A^+A = 0$ and thus $(A^+A)^*(BB^+)^* = 0$ which implies $A^+ABB^+ = 0$. Hence $AA^+ABB^+B = AB = 0$.

P28) Let A be an m by n matrix, and x any n-component column vector. Then

$$Ax = 0 \text{ if and only if } x^*A^+ = 0.$$

Proof: Assume $Ax = 0$. Then $A^+Ax = 0$, which implies that $x^*(A^+A)^* = 0$ or $x^*A^+A = 0$. Multiplying by A^+ we bet $x^*A^+AA^+ = x^*A^+ = 0$. Conversely, if $x^*A^+ = 0$, then $x^*A^+A = 0$ or $x^*(A^+A)^* = 0$ which implies that $A^+Ax = 0$. Multiplying by A yields $AA^+Ax = Ax = 0$.

P29) If U and V are unitary, $(UAV)^{+} = V^{*}A^{+}U^{*}$.

Proof: By direct substitution of $V^{*}A^{+}U^{*}$ in the four defining equations for $(UAV)^{+}$ we obtain the desired result.

P30) If P is hermitian and idempotent, $(PA)^{+} = Q^{+}P$ whenever either $PQ = Q$ or P commutes with Q, $Q^{+}Q$ and QQ^{+}.

Proof: By direct substitution into Theorem 2.1.

P31) Let C be a square matrix in Jordan canonical form. $(C - uI)(C - uI)^{+}x = 0$ if and only if x is an eigenvector of C^{*} corresponding to the eigenvalue $\bar{u}$.

Proof: Assume $(C - uI)(C - uI)^{+}x = 0$. Since $X = R(C - uI) \oplus N[(C - uI)^{+}]$ we can write $x = x_1 + x_2$ where $x_1 \varepsilon R(C - uI)$ and $x_2 \varepsilon N[(C - uI)^{+}]$. Then $(C - uI)(C - uI)^{+}(x_1 + x_2) = x_1 = 0$ so that $x = x_2$. Now $N[(C - uI)^{+}] = N[(C - uI)^{*}] = N[C - \bar{u}I]$.

Hence $(C^{*} - \bar{u}I)x = 0$ or $C^{*}x = \bar{u}x$.

Assuming the converse, we have that $C^{*}x = \bar{u}x$ or $(C^{*} - \bar{u}I)x = 0$. This implies that $x \varepsilon N[(C^{*} - \bar{u}I)] = N[(C - uI)^{+}]$. Hence

$$(C - uI)(C - uI)^{+}x = 0 .$$

P32) Let $\{A_i\}$, $i = 1, 2, \ldots, k$ be arbitrary m by n matrices. Then

$$A_i - A_i \left(\sum_{j=1}^{k} A_j^T A_j \right)^+ \left(\sum_{j=1}^{k} A_j^T A_j \right) = 0$$

for all $i = 1, 2, \ldots, k$.

<u>Proof</u>: Let $S = \sum_{j=1}^{k} A_j^T A_j$ and consider

$A_i^* A_i - S^+ S A_i^* A_i$. Since S is normal, $S^+ S = S S^+$ and thus we can write

$$A_i^* A_i - S^+ S\, A_i^* A_i = A_i^* A_i - S S^+ A_i^* A_i \ .$$

Now SS^+ is an orthogonal projection on the range space of S which contains the range space of $A_i^* A_i$. Hence, $SS^+ A_i^* A_i = A_i^* A_i$, so that $A_i^* A_i - S^+ S\, A_i^* A_i = 0$. Taking the conjugate transpose of both sides of this equation and using the fact that if $A^* B = 0$ with the columns of B in $R(A)$, then $B = 0$, the result follows.

P33) Let A be an m by n matrix, $m \geq n$, then

$$|\lambda I_m - AA^+| = \lambda^{m-n} |\lambda I_n - A^+ A| .$$

Proof: First we give a simple proof for the case $m = n$. Let the zeros of $|\lambda I_n - A^+A|$ be distinct, say $\lambda_1, \ldots, \lambda_n$. If $\lambda = 0$ is an eigenvalue of A^+A it is an eigenvalue of AA^+, since $|A^+A| = |AA^+|$. For $\lambda_1 \neq 0$, $x_1 \neq 0$, it follows from $A^+Ax_1 = \lambda_1 x_1$ that $Ax_1 \neq 0$. Hence $AA^+Ax_1 = \lambda_1 Ax_1$, so that λ_1 is an eigenvalue of AA^+. Thus every eigenvalue of A^+A is an eigenvalue of AA^+, and the result holds for $m = n$. If multiple zeros of $|\lambda I_n - A^+A|$ exist, one need only add small quantities to the elements of A and A^+ such that the zeros of $|\lambda I_n - A^+A|$ separate and become distinct. Thus

$$|\lambda I_n - A(\varepsilon)A^+(\varepsilon)| = |\lambda I_n - A^+(\varepsilon)A(\varepsilon)| \quad \text{with} \quad A(o) = A ,$$

$A^+(o) = A^+$, and ε represents a set of small elements. From continuity considerations the result holds for $m = n$.

We consider next the case $m > n$, or $m = n + p$, $p > 0$. Let M be the augmented matrix $M = (A, \phi_1)$ with ϕ_1 the $m \times p$ null matrix, and let

$$N = \begin{bmatrix} A^+ \\ \\ \phi_2 \end{bmatrix} ,$$

with ϕ_2 the $p \times m$ null matrix, $\phi_1 = \phi_2^T$. Thus, M and N are square matrices of order m. It follows that

$$|\lambda I_m - MN| = |\lambda I_m - NM|. \tag{1}$$

One notes that

$$MN = (A,\phi_1)\begin{pmatrix} A^+ \\ \phi_2 \end{pmatrix} = AA^+, \quad NM = \begin{pmatrix} A^+ \\ \phi_2 \end{pmatrix}(A,\phi_1) = \begin{pmatrix} A^+A & \phi_3 \\ \phi_4 & \phi_5 \end{pmatrix} .$$

Hence (1) becomes

$$|\lambda I_m - AA^+| = \begin{vmatrix} \lambda I_n - A^+A & \phi_3 \\ \phi_4 & \lambda I_p \end{vmatrix} = \lambda^{m-n}|\lambda I_n - A^+A|$$

which concludes the proof.

It should be noted that P33 holds for the more general case where A^+ is replaced by any $n \times m$ matrix.

P34) The following conditions are each necessary and sufficient for $(AB)^+ = B^+A^+$.

1) $A^+ABB^*A^* = BB^*A^*$ and $BB^+A^*AB = A^*AB$

2) Both A^+ABB^* and A^*ABB^+ are hermitian

3) $A^+ABB^*A^*ABB^+ = BB^*A^*A$

4) $A_2^*A_1 = 0$, $B_2B_1^* = 0$ where

$$A_1 = ABB^+ , \; A_2 = A - A_1 ,$$

$$B_1 = A^+AB , \; B_2 = B - B_1 .$$

If A and B are otherwise arbitrary matrices such that AB is defined, $(AB)^+ = B^+A^+$ if and only if both the equations

$$A^+ABB^*A^* = BB^*A^* \tag{2}$$

and

$$BB^+A^*AB = A^*AB \tag{3}$$

are satisfied.

<u>Proof</u>: Multiplying $A^+ABB^*A^* = BB^*A^*$ on the right by $(AB)^{*+}$ and using $C^+CC^* = C^*CC^+ = C^*$, and $CC^*C^{*+} = C^{*+}C^*C = C$, in the form

$$(AB)(AB)^*(AB)^{*+} = AB ,$$

gives

$$B^+A^+AB = (AB)^*(AB)^{*+} = (AB)^+(AB) . \tag{4}$$

Similarly, taking transposes of both sides of (3) gives

$$B^*A^*ABB^+ = B^*A^*A ,$$

and then multiplying on the right by A^+ and on the left by $(AB)^{*+}$ and using $C^+CC^* = C^*CC^+ = C^*$, and $CC^*C^{*+} = C^{*+}C^*C = C$, leads to the equation

$$ABB^+A^+ = AB(AB)^+ . \tag{6}$$

Recognizing that $(AB)(AB)^+$ and $(AB)^+(AB)$ are the orthogonal projectors on the range spaces $R(AB)$ and $R((AB)^*)$, respectively,

(4) and (6) express the fact that B^+A^+ is the generalized inverse of AB, as defined by Moore [66].

Conversely, $(AB)^+ = B^+A^+$ implies

$$B^*A^* = B^+A^+ABB^*A^* .$$

multiplying on the left by ABB^*B and using $B^*BB^+ = B^*$ gives

$$ABB^* (I - A^+A) BB^*A^* = \theta,$$

where θ denotes a null matrix. As the left member is Hermitian and $I - A^+A$ is idempotent, it follows that

$$(I - A^+A) BB^*A^* = \theta,$$

which is equivalent to (2). In an analogous manner, (3) is obtained.

P35) $(AB)^+ = B^+A^+$ if and only if both A^+ABB^* and A^*ABB^+ are Hermitian.

<u>Proof</u>: If A^+ABB^* is Hermitian, we have

$$A^+ABB^* = BB^*A^+A ,$$

and multiplication on the right by A^* gives (2). Conversely, multiplication of (2) on the right by A^{*+} gives

$$A^+ABB^*A^+A = BB^*A^+A . \tag{7}$$

Since the left member of (7) is Hermitian, the right member is also.

In a similar fashion it can be shown that (3) is equivalent to the statement that A^*ABB^+ is Hermitian.

It will be noted that an equivalent statement to the condition in P35) is that A^+A and BB^* commute and also A^*A and BB^+ commute.

P36). $(AB)^+ = B^+A^+$ if and only if

$$A^+ABB^*A^*ABB^+ = BB^*A^*A \ . \tag{8}$$

<u>Proof</u>.: Multiplying (8) on the left by A^+A gives

$$A^+ABB^*A^*ABB^+ = A^+ABB^*A^*A \ . \tag{9}$$

Combining (8) and (9) gives

$$A^+ABB^*A^*A = BB^*A^*A \ ,$$

and multiplication on the right by A^+ gives (2). An analogous process leads to (5), which is equivalent to (3).

On the other hand, if (2) and (3) hold, multiplying (2) on the right by A and then using (5) to transform the left member gives (8).

Equations (2) and (3) have a simple interpretation in terms of range spaces. They assert, respectively, that $R(A^*)$ is an invariant space of BB^* and that $R(B)$ is an invariant space of A^*A. In some particular cases this interpretation leads to a characterization of those matrices B that satisfy $(AB^+ = B^+A^+$ for a given A. For example, if A is of full column rank, $A^+A = I$ and (2) is immediately satisfied. Then (3) holds if and only if B is a null matrix or $R(B)$ is the space spanned by some set of eigenvectors of A^*A .

P37). $(AB)^+ = B^+A^+$ if and only if both the equations

$$A^+AB = B(AB)^+AB \tag{10}$$

and

$$BB^+A^* = A^* AB(AB)^+ \tag{11}$$

are satisfied.

Proof: Multiplication of (2) on the right by $(AB)^{*+}$ gives (10), and conversely multiplication of (10) on the right by $(AB)^*$ gives (2). Similarly it can be shown that (11) is equivalent to (5).

P38). A necessary condition for $(AB)^+ = B^+A^+$ is that A^+A and BB^+ commute.

Proof: Substitution of B^+A^+ for $(AB)^+$ in (10) and multiplication on the right by B^+ gives

$$A^+ABB^+ = BB^+ A^+ ABB^+ .$$

As the right member is Hermitian, the conclusion follows.

That the condition of P38) is not sufficient is clear from the example:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (AB)^+ = (0 \quad 1) \quad B^+A^+ = (-1 \quad 1) .$$

As A is nonsingular, $A^+A = A^{-1}A = I$, and the condition is fulfilled.

It is easily seen that the commutativity of A^+A and BB^+ is equivalent to either of the conditions

$$A^+ABB^+ A^* = BB^+ A^*$$

and

$$BB^+ A^+ AB = A^+ AB .$$

These equations can be interpreted as asserting that $R(A^*)$ is the direct sum of a subspace of $R(B)$ and a space orthogonal to $R(B)$ and that $R(B)$ is the direct sum of a subspace of $R(A^*)$ and a space orthogonal to $R(A^*)$. These observations reveal something about the structure of matrices A and B that satisfy $(AB)^+ = B^+A^+$. It is easily seen that (2) and (3) are equivalent to the following two equations:

$$(I - A^+A)\, BB^*\, A^+\, A = \theta . \tag{12}$$

$$(I - BB^+)A^*ABB^+ = \theta . \tag{13}$$

Equation (12) shows that if B is resolved into the two component matrices,

$$B_1 = A^+ AB \qquad B_2 = (I - A^+ A)B ,$$

then not only do we have $B_1^*B_2 = \theta$ as expected, but also $B_2B_1^* = \theta$. Similar remarks apply to the resolution of A^* into

$$A_1^* = BB^+ A^* \qquad A_2^* = (I - BB^+)A^* .$$

3.2 Representations for the Pseudoinverse of a Partitioned Matrix

Let $A = (A_{k-1}\ a_k)$ where a_k is the k^{th} column of A and A_{k-1} is the submatrix of A consisting of the first $k-1$ columns.

Let $d_k = A^+_{k-1}a_k$ and $c_k = a_k - A_{k-1}d_k$. Then

$$A^+ = \begin{bmatrix} A^+_{k-1} - d_k b_k \\ \\ b_k \end{bmatrix} \tag{1}$$

where

$$b_k = \begin{cases} c_k^+, & \text{if } c_k \neq 0 \\ \\ (1 + d_k^* d_k)^{-1}\ d_k^* A_{k-1}^+, & \text{if } c_k = 0 \end{cases} \tag{2}$$

It is computational to establish that this form of A^+ satisfies the four defining equations in theorem 2.1.

The form of A^+ will now be extended to obtain representations for the pseudoinverse of matrices $A = (U, V)$. We begin by combining (1) and (2) into a single expression.

Since c_k is a single column vector, $c_k \neq 0$ implies $c_k^+ = (c_k^* c_k)^{-1} c_k^*$ and thus $c_k^+ c_k = 1$. Further, $c_k = 0$ implies $c_k^+ = 0$ and $c_k^+ c_k = 0$. Then we can rewrite b_k as

$$b_k = c_k^+ + (1 - c_k^+ c_k)\ (1 + d_k^* d_k)^{-1}\ d_k^* A_{k-1}^+ \tag{3}$$

and obtain a single expression for the cases $c_k = 0$ and $c_k \neq 0$ in (2).

Combining (1) and (3) then gives

$$A^+ = \begin{bmatrix} A_{k-1}^+ - A_{k-1}^+ a_k c_k^+ - A_{k-1}^+ a_k (1 - c_k^+ c_k)\ k_1 a_k^* A_{k-1}^{+*} A_{k-1}^+ \\ \\ c_k^+ + (1 - c_k^+ c_k) k_1 a_k^* A_{k-1}^{+*} A_{k-1}^+ \end{bmatrix} \qquad (4)$$

where k_1 designates the quantity $(1 + d_k^* d_k)^{-1}$ and $a_k^* A_{k-1}^{+*}$ is utilized in place of d_k^*. The expression in (4) exhibits the structure of the representations for the generalized inverse of matrices $A = [U, V]$.

Consider an arbitrary matrix $A = [U, V]$, where U and V have ℓ and $k - \ell$ columns, respectively, $0 \leqq \ell \leqq k$. Corresponding to c_k and k_1 in (4) let $C = (I - UU^+)V$ and $K_1 = (I + V^* U^{+*} U^+ V)^{-1}$, and let

$$X_1 = \begin{bmatrix} U^+ - U^+ V C^+ - U^+ V (I - C^+ C) K_1 V^* U^{+*} U^+ \\ \\ C^+ + (I - C^+ C) K_1 V^* U^{+*} U^+ \end{bmatrix} \qquad (5)$$

Then we have

<u>Theorem 3.2</u> A necessary and sufficient condition that $X_1 = A^+$ is that the matrices $C^+ C$ and $V^* U^{+*} U^+ V$ commute.

<u>Proof</u>: It will be shown that A and X_1 satisfy the defining equations in theorem 2.1, where the commutativity of $C^+ C$ and $V^* U^{+*} U^+ V$ is utilized in order to conclude that $(X_1 A)^* = X_1 A$.

Using the definition of C and the relation $CC^+C = C$ to simplify the resulting expression, block multiplication gives

$$AX_1 = UU^+ + CC^+ . \tag{6}$$

Thus, since both UU^+ and CC^+ are hermitian, $(AX_1)^* = AX_1$.

Now $U^+C = (U^+ - U^+UU^+)V = 0$ implies $C^+U = 0$, by (P1) and (P27). Whence

$$UU^+V + CC^+V = UU^+V + CC^+C = UU^+V + C = V ,$$

and the product $AX_1A = (AX_1)A$ becomes

$$AX_1A = [(UU^+ + CC^+)U, (UU^+ + CC^+)V] = [U, V] = A .$$

Similarly, $X_1AX_1 = X_1(AX_1)$ reduces to

$$X_1AX_1 = \begin{bmatrix} U^+ - U^+VC^+ - U^+V(I - C^+C)K_1V^*U^{+*}U^+ \\ \\ C^+ + (I - C^+C)K_1V^*U^{+*}U^+ \end{bmatrix} = X_1 ,$$

since $U^+(UU^+ + CC^+) = U^+$ and $C^+(UU^+ + CC^+) = C^+$.

Finally, with $C^+U = 0$, $C^+V = C^+C$ and $U^{+*}U^+U = U^{+*}(U^+U)^* = (U^+UU^+)^* = U^{+*}$, the product X_1A becomes

$$X_1A = \begin{bmatrix} U^+U - U^+V(I - C^+C)K_1V^*U^{+*} & U^+V(I - C^+C)K_1 \\ \\ (I - C^+C)K_1V^*U^{+*} & I - (I - C^+C)K_1 \end{bmatrix} \tag{7}$$

where $K_1V^*U^{+*}U^+V = I - K_1$ by definition of K_1. Suppose now that C^+C and $V^*U^{+*}U^+V$ commute. Then

$$(I - C^+C)\ (I + V^*U^{+*}U^+V) = (I + V^*U^{+*}U^+V)\ (I - C^+C)\ ,$$

and so

$$K_1(I - C^+C) = (I - C^+C)K_1\ . \tag{8}$$

Since both K_1 and $I - C^+C$ are hermitian, this implies

$$[K_1(I - C^+C)]^* = K_1(I - C^+C)\ , \tag{9}$$

and it follows in (7) that $(X_1A)^* = X_1A$.

Thus we have shown that A and X_1 satisfy the relations $AX_1A = A$, $X_1AX_1 = X_1$, $(AX_1)^* = AX_1$ and $(X_1A)^* = S_1A$, provided C^+C and $V^*U^{+*}U^+V$ commute, and so $X_1 = A^+$.

Conversely, if $X_1 = A^+$, then $(X_1A)^* = X_1A$, and (9) follows at once from (7). Hence (8) holds, and C^+C commutes with $V^*U^{+*}U^+V$.

The existence of matrices $A = [U, V]$ for which C^+C and $V^*U^{+*}U^+V$ do not commute can be shown by simple examples. Consequently, X_1 does not provide the most general form for A^+. Before considering the general form, however, we will establish four corollaries to theorem 3.2. Assume A has the form $A = [U, V]$, and again let $C = (I - UU^+)V$ and $K_1 = (I + V^*U^{+*}U^+V)^{-1}$.

<u>Corollary 3.1</u>

$$A^+ = \begin{bmatrix} U^+ - U^+VK_1V^*U^{+*}U^+ \\ \\ C^+ + K_1V^*U^{+*}U^+ \end{bmatrix} \tag{10}$$

if and only if $C^+CV^*U^{+*}U^+V = 0$.

Proof: If $C^+CV^*U^{+*}U^+V = 0$, then $V^*U^{+*}U^+V$ and C^+C commute, and $X_1 = A^+$. Also, $C^+CV^*U^{+*}U^+V = 0$ implies $C^{+*}V^*U^{+*}U^+VC^+ = 0$ and thus

$$U^+VC^+ = 0. \tag{11}$$

Whereupon X_1 in (5) reduces at once to the right hand side of (10).

Conversely, if A^+ has the form given in (10), then it follows from the equation $AA^+A = A$ that

$$UU^+V - UU^+VK_1V^*U^{+*}U^+V + VC^+V + VK_1V^*U^{+*}U^+V = V.$$

Using the relations $K_1 - I - K_1V^*U^{+*}U^+V$ and $C^+V = C^+C$, the definition of C now gives

$$C(I - K_1) + VC^+C = C$$

and so

$$C^+C\ (I - K_1) = 0\ .$$

Hence

$$C^+C\ (I + V^*U^{+*}U^+V) = C^+C$$

and

$$C^+CV^*U^{+*}U^+V = 0\ .$$

Note in Corollary 3.1 that $C^+CV^*U^{+*}U^+V = \theta$ is equivalent to the condition $VC^+V = C$. If $C^+CV^*U^{+*}U^+V = \theta$, then we have, using (11),

$$VC^+V = VC^+C = VC^+C - UU^+VC^+C = CC^+C = C\ .$$

Conversely, if $VC^+V = C$, then

$$V^*U^{+*}U^+VC^+C = V^*U^{+*}U^+VC^+V = V^*U^{+*}U^+C = \theta ,$$

and thus $C^+CV^*U^{+*}U^+V = \theta$.

For the special case in which $C = \theta$, Corollary 3.1 reduces to

Corollary 3.2

$$A^+ = \begin{bmatrix} U^+ - U^+VK_1V^*U^{+*}U^+ \\ \\ K_1V^*U^{+*}U^+ \end{bmatrix} \tag{12}$$

if and only if $C = 0$.

Proof: That (10) reduces to (12) when $C = 0$ is obvious. Conversely, if (12) holds, then $AA^+A = A$ implies

$$UU^+V - UU^+VK_1V^*U^{+*}U^+V + VK_1V^*U^{+*}U^{+*}U^+V = V ,$$

which reduces to $C(I - K_1) = C$. Hence $CK_1 = 0$ and so $C = 0$.

Suppose now that $C^+CV^*U^{+*}U^+V = V^*U^{+*}U^+V$. Then again $V^*U^{+*}U^+V$ and C^+C commute and we obtain two more special cases of theorem 3.1,

Corollary 3.3

$$A^+ = \begin{bmatrix} U^+ - U^+VC^+ \\ \\ C^+ \end{bmatrix} \tag{13}$$

if and only if

$$C^+CV^*U^{+*}U^+V \;=\; V^*U^{+*}U^+V\;. \tag{14}$$

Proof: From P4), $B^*BB^+ = B^*$ for every matrix B. Hence, taking $B = U^+V$, the relation $C^+CV^*U^{+*}U^+V \;=\; V^*U^{+*}U^+V$ implies that

$$C^+CV^*U^{+*} \;=\; V^*U^{+*}\;, \tag{15}$$

and using (8) gives

$$(I - C^+C)K_1V^*U^{+*}U^+ \;=\; 0\;.$$

Whereupon X_1 reduces to (13).

Conversely, if A^+ has the form given in (13), then A^+A becomes

$$A^+A \;=\; \begin{bmatrix} U^+U & U^+V(I - C^+C) \\ & \\ \theta & C^+C \end{bmatrix}$$

and $(A^+A)^* \;=\; A^+A$ implies $(I - C^+C)V^*U^{+*} \;=\; \theta$. This gives (15), from which the converse follows.

Analogous to the equivalence between the conditions $C^+CV^*U^{+*}U^+V = \theta$ and $VC^+V = C$ noted above, it is easily seen that $C^+CV^*U^{+*}U^+V = V^*U^{+*}U^+V$ is equivalent to having $VC^+V = V$. If $C^+CV^*U^{+*}U^+V = V^*U^{+*}U^+V$, then it follows from (15) and the definition of C that

$$V^* - C^+CV^*U^{+*}U^* \;=\; V^* - V^*U^{+*}U^*$$

or

$$V^* - C^+C(V^* - C^*) \;=\; C^*\;.$$

Thus $V^* - C^+CV^* = \theta$ and so $V = VC^+C \;=\; VC^+V$.

Conversely, if $VC^+V = V$, then

$$V^*U^{+*}U^+V = V^*U^{+*}U^+VC^+V = V^*U^{+*}U^+VC^+C ,$$

which implies

$$C^+CV^*U^{+*}U^+V = V^*U^{+*}U^+V .$$

Note, in particular, that whenever C has full column rank we have $C^+ = (C^*C)^{-1}C^*$, by (P12), and thus $C^+C = I$. Hence $VC^+V = VC^+C = V$, and A^+ has the form given in (13). Clearly, this is the case in the form for A^+, (1), when $c_k \neq \theta$ and $b_k = c_k^+$ in (2). On the other hand, when $c_k = \theta$, Corollary 3.2 is applicable, and the form for A^+ with $b_k = (1 + d_k^*d_k)^{-1}d_k^*A_{k-1}^+$ follows directly from (12).

For the special case in which $C = V$, $VC^+V = VV^+V = V$, and Corollary 3.3 reduces to

<u>Corollary 3.4</u>

$$A^+ = \begin{bmatrix} U^+ \\ \\ V^+ \end{bmatrix} \tag{16}$$

if and only if $C = V$.

<u>Proof</u>: If $C = V$, then $UU^+V = \theta$ and

$$U^+VC^+ = U^+UU^+VC^+ = \theta$$

in the form for A^+, (13), which gives (16).

Conversely, if (16) holds, then it follows from the relation $AA^+A = A$ that

$$UU^+V + VV^+V = V.$$

Hence $UU^+V = \theta$ and $C = V$.

Let us now consider general forms for A^+ in which it is not required that C^+C and $V^*U^{+*}U^+V$ commute. Let $\tilde{C}$ designate the expression

$$\tilde{C} = (I - VV^+)\, U$$

obtained by interchanging the roles of U and V in $C = (I - UU^+)V$. Also, let K and $\tilde{K}$ designate the dual expressions defined by

$$K = [I + (I - C^+C)V^*U^{+*}U^+V(I - C^+C)]^{-1} \tag{17}$$

$$\tilde{K} = [I + (I - \tilde{C}^+\tilde{C})U^*V^{+*}V^+U(I - \tilde{C}^+\tilde{C})]^{-1} . \tag{18}$$

(Note that both K and $\tilde{K}$, inverses of positive definite matrices, exist for every U and V.) Then we have

<u>Theorem 3.3</u>. The generalized inverse of any matrix, $A = [U,V]$ can be written in the following equivalent forms.

(a) $$A^+ = \begin{bmatrix} U^+ - U^+VC^+ - U^+V(I - C^+C)KV^*U^{+*}U^+(I - VC^+) \\ \\ C^+ + (I - C^+C)KV^*U^{+*}U^+(I - VC^+) \end{bmatrix}$$

(b) $$A^+ = \begin{bmatrix} U^+ - U^+VC^+ - U^+V(I - C^+C)KV^*U^{+*}U^+(I - VC^+) \\ \\ V^+ = V^+U\tilde{C}^+ - V^+U(I - \tilde{C}^+\tilde{C})\tilde{K}U^*V^{+*}V^+(I - U\tilde{C}^+) \end{bmatrix}$$

(c) $$A^{+} = \begin{bmatrix} \tilde{C}^{+} + (I - \tilde{C}^{+}\tilde{C})\tilde{K}U^{*}V^{+*}V^{+}(I - U\tilde{C}^{+}) \\ \\ C^{+} + (I - C^{+}C)KV^{*}U^{+*}U^{+}(I - VC^{+}) \end{bmatrix} .$$

<u>Proof</u>: Let X_0 designate the matrix

$$X_0 = \begin{bmatrix} U^{+} - U^{+}VC^{+} - U^{+}V(I - C^{+}C)L \\ \\ C^{+} + (I - C^{+}C)L \end{bmatrix} \tag{19}$$

obtained from X_1 in (5) by replacing the quantity $K_1V^{*}U^{+*}U^{+}$ by an arbitrary matrix L, of the same size. Then it follows immediately, using block multiplication, the definition of C, and the relation $C(I - C^{+}C) = \theta$, that $AX_0 = UU^{+} + CC^{+} = AX_1$, and so we have $(AX_0)^{*} = AX_0$ and $AX_0A = A$ from the proof of Theorem 3.1.

Now forming $X_0AX_0 = X_0(AX_0)$, it is clear that $X_0AX_0 = X_0$ provided L satisfies

$$L(UU^{+} + CC^{+}) = L . \tag{20}$$

Similarly, forming X_0A gives

$$X_0A = \begin{bmatrix} U^{+}U - U^{+}V(I - C^{+}C)LU & U^{+}V(I - C^{+}C)(I - LV) \\ \\ (I - C^{+}C)LU & C^{+}C + (I - C^{+}C)LV \end{bmatrix}$$

upon simplification of the submatrices, and it follows that $(X_0A)^{*} = X_0A$ provided L satisfies

$$[U^+V(I - C^+C)(I - LV)]^* = (I - C^+C)LU \tag{21}$$

and also that both $U^+V(I - C^+C)LU$ and $(I - C^+C)LV$ are hermitian. We will now show that the expression

$$L = KV^*U^{+*}U^+(I - VC^+) , \tag{22}$$

with K as defined in (17), satisfies these conditions.

Since $U^+(I - VC^+)UU^+ = U^+$ and $U^+(I - VC^+)CC^+ = -U^+VC^+$, then L in (22) satisfies (20). Next observe that since $I - C^+C$ is idempotent, it commutes with the matrix $I + (I - C^+C)V^*U^{+*}U^+V(I - C^+C)$, and thus with K. Whereupon, with both $I - C^+C$ and K hermitian,

$$[(I - C^+C)K]^* = (I - C^+C)K,$$

and so

$$U^+V(I - C^+C)LU = U^+V(I - C^+C)KV^*U^{+*}$$

is hermitian. Moreover, we have

$$(I - C^+C)LV = (I - C^+C)K(I - C^+C)V^*U^{+*}U^+V(I - C^+C) ,$$

or

$$(I = C^+C)LV = (I - C^+C)(I - K) ,$$

which implies that $(I - C^+C)LV$ is hermitian. Finally, since

$$U^+V(I - C^+C)(I - LV) = U^+V(I - C^+C)K = [(I - C^+C)KV^*U^{+*}]^*$$

and

$$(I - C^+C)LU = (I - C^+C)KV^*U^{+*} ,$$

then (21) holds for this choice of L.

Thus it has been shown that X_0 and A satisfy the relations $AX_0A = A$, $X_0AX_0 = X_0$, $(AX_0)^* = AX_0$, and $(X_0A)^* = X_0A$, provided L has the form given in (22), and so $X_0 = A^+$. The form for A^+ in (a) is obtained by replacing L in (19) by the expression in (22).

The forms for A^+ in (b) and (c) are now easily established. Let $\tilde{A}$ designate the matrix $\tilde{A} = [V, U]$. Then it follows from (a) that $\tilde{A}^+$ can be written as

$$\tilde{A}^+ = \begin{bmatrix} V^+ - V^+U\tilde{C}^+ - V^+U(I - \tilde{C}^+\tilde{C})\tilde{K}U^*V^{+*}V^+ (I - U\tilde{C}^+) \\ \\ \tilde{C}^+ + (I - \tilde{C}^+\tilde{C})\tilde{K}U^*V^{+*}V^+ (I - U\tilde{C}^+) \end{bmatrix} \tag{23}$$

where $\tilde{C}$ and $\tilde{K}$ are the dual expressions obtained from C and K by interchanging the roles of U and V. Since A and $\tilde{A}$ differ only by the order in which columns are written, there is a unitary permutation matrix P, say, such that $A = \tilde{A}P$. Then we have $A^+ = P^*\tilde{A}^+$, by (P29). Now P as a right multiplier permutes columns of $\tilde{A}$, and P^* as a left multiplier permutes rows of $\tilde{A}^+$ in the same order, and it follows from (23) that A^+ can be written in the form

$$A^+ = \begin{bmatrix} \tilde{C}^+ + (I - \tilde{C}^+\tilde{C})\tilde{K}U^*V^{+*}V^+ (I - U\tilde{C}^+) \\ \\ V^+ - V^+U\tilde{C}^+ = V^+U(I - \tilde{C}^+\tilde{C})\tilde{K}U^*V^{+*}V^+ (I - U\tilde{C}^+) \end{bmatrix} \tag{24}$$

But A^+ is unique. The forms for A^+ in (b) and (c) are obtained now by equating the corresponding expressions for submatrices in (a) and (24).

It also follows from the symmetry exhibited by the expressions for A^+ in Theorem 3.3 (a) and (24) that Theorem 3.2 and each of its corollaries has a corresponding dual form in which the roles of U and V are interchanged.

Consider an arbitrary matrix $A = [U, V]$, and assume A^+ is known. Partition A^+ as $A^+ = \binom{G}{H}$ where G and H have the size of U^* and V^*, respectively. Then Theorem 3.4 provides an expression for U^+ in terms of $G, H,$ and related matrices.

Theorem 3.4.

$$U^+ = G[I + V(I - HV)^+H]$$

$$\cdot\{I - [H - (I - HV)(I - HV)^+H]^+ [H - (I - HV)(I - HV)^+H]\}.$$

Proof: We know from the expression in Theorem 3.3 (a) that

$$G = U^+ - U^+VC^+ - U^+V(I - C^+C)KV^*U^{+*}U^+(I - VC^+)$$

and

$$H = C^+ + (I - C^+C)KV^*U^{+*}U^+(I - VC^+)$$

in the partition of A^+ corresponding to the partition $A = [U, V]$. Then it follows using the relations employed in the proof of Theorem 3.2 (a) that

$$GV = U^+V(I - C^+C)K \tag{25}$$

and

$$I - HV = (I - C^+C)K. \tag{26}$$

Further, $I - C^+C$ is idempotent and hermitian, and commutes with K. Therefore, since K is nonsingular, we have

$$(I - HV)^{+} = K^{-1} (I - C^{+}C) ,$$

by (P30), which combined with (25) and (26) to give

$$GV (I - HV)^{+} = U^{+}V (I - C^{+}C) \tag{27}$$

and

$$(I - HV)(I - HV)^{+} = I - C^{+}C. \tag{28}$$

Now since $C^{+}CC^{+} = C^{+}$,

$$GV(I - HV)^{+}H = U^{+}V(I - C^{+}C)KV^{*}U^{+*}U^{+}(I - VC^{+}) ,$$

and so

$$G[I + V(I - HV)^{+}H] = U^{+} - U^{+}VC^{+} . \tag{29}$$

Moreover

$$(I - HV)(I - HV)^{+}H = (I - C^{+}C)KV^{*}U^{+*}U^{+}(I - VC^{+}) ,$$

and thus

$$H - (I - HV)(I - HV)^{+}H = C^{+} . \tag{30}$$

Finally, since $U^{+}C = \theta$, we have

$$U^{+} = G[I + V(I - HV)^{+}H](I - CC^{+})$$

from (29), which combines with (30) and the relation $C^{++} = C$ to give the stated form for U^{+}.

The following corollaries provide special forms for U^{+} corresponding to the forms for A^{+} in corollaries to Theorem 3.1. This correspondence is apparent by observing that the relations satisfied by V and C are simply the alternative statements of the conditions on $C^{+}C$ and $V^{*}U^{+*}U^{+}V$ which were noted above.

Corollary 3.5. $U^{+} = G[I + V(I - HV)^{+}H]$ if and only if $VC^{+}V = C$.

Proof: It follows from (29) that $U^{+} = G[I + V(I - HV)^{+}H]$ if and only if $U^{+}VC^{+} = \theta$. But this implies $VC^{+}V = C$, and conversely.

Corollary 3.6. $U^{+} = G(I - H^{+}H)$ if and only if $VC^{+}V = V$.

Proof: From Corollary 3.3 we have $G = U^{+} - U^{+}VC^{+}$ and $H = C^{+}$ if $VC^{+}V = V$. Hence $G[I + V(I - HV)^{+}H] = G$ in (29) and $H - (I - HV)(I - HV)^{+}H = H$ in (30), and the general form for U^{+} in Theorem 3.4 reduces immediately to the above expression.

Conversely, since we can write the general form for G as $G = U^{+}(I - VH)$, then

$$U^{+} = U^{+}(I - VH)(I - H^{+}H) = U^{+}(I - H^{+}H)$$

if $U^{+} = G(I - H^{+}H)$. This gives $U^{+}H^{+}H = \theta$ and so

$$UU^{+}H^{*} = UU^{+}H^{*}H^{+*}H^{*} = UU^{+}H^{+}HH^{*} = \theta,$$

Then, $HUU^{+} = \theta$, since UU^{+} is hermitian, and $\theta = HUU^{+} =$

$$C^{+}UU^{+} + (I - C^{+}C)KV^{*}U^{+*}U^{+}(I - VC^{+})UU^{+} = (I - C^{+}C)KV^{*}U^{+*}U^{+},$$

which implies $G = U^{+} - U^{+}VC^{+}$ and $H = C^{+}$. Whence $C^{+}CV^{*}U^{+*}U^{+}V = V^{*}U^{+*}U^{+}V$, by Corollary 3.3, and thus $VC^{+}V = V$.

Corollary 3.7. $U^{+} = G$ if and only if $C = V$.

Proof: That $C = V$ implies $U^{+} = G$ follows directly from Corollary 3.4. Conversely, if $U^{+} = G$, we have

$$U^{+}VC^{+} + U^{+}V(I - C^{+}C)KV^{*}U^{+*}U^{+}(I - VC^{+}) = \theta, \tag{31}$$

by definition of G. Multiplying (30) on the right by VC^+ then gives $U^+VC^+ = \theta$, and thus

$$U^+VK_1V^*U^{+*}U^+ = \theta,$$

where $K_1 = (I + V^*U^{+*}U^+V)^{-1}$. Therefore $U\,V(I - K_1) = \theta$ and $U^+VV^*U^{+*}U^+V = \theta$, from which it follows that $U^+V = \theta$ and $C = V$.

Observe in Corollary 3.5 that $VC^+V = C$ if $C = \theta$. In this case $H = K_1V^*U^{+*}U^+$, and $I - HV = K_1$ is nonsingular, by definition of K_1. Conversely, if $I - HV$ is nonsingular, then $C^+C = \theta$, by (28) which implies $C = \theta$. Thus is follows that $I - HV$ nonsingular is a necessary and sufficient condition that A^+ has the form given in Corollary 3.2. Since we can have $VC^+V = C$ but $C \neq \theta$, $I - HV$ nonsingular is only a sufficient condition that U^+ has the form given in Corollary 3.5. In Corollaries 3.6 and 3.7, however, the necessary and sufficient conditions that U^+ has the simplified forms can be restated in terms of V and H. This gives Corollaries 3.6 (a) and 3.7 (a).

Corollary 3.6(a). $U^+ = G(I - H^+H)$ if and only if HV is idempotent.

Proof: From Corollary 3.6 it follows that we only need to show that HV idempotent implies $VC^+V = V$, and conversely.

If $VC^+V = V$, then $H = C^+$, by Corollary 3.3 and

$$(HV)^2 = C^+VC^+V = C^+V = HV.$$

Conversely, HV idempotent implies that $I - HV$ is idempotent. Therefore, since we also have $I - HV = (I - C^+C)K$ hermitian, from the proof of Theorem 3.3(a), $(I - HV)^+ = I - HV$. Now $(I - HV)^+ = K^{-1}(I - C^+C)$, and thus

$$K^{-1}(I - C^+C) = (I - HV)(I - HV)^+ = I - C^+C,$$

by (28). This gives

$$(I - C^+C)V^*U^{+*}U^+V(I - C^+C) = \theta,$$

by definition of K. Then $(I - C^+C)V^*U^{+*} = \theta$, and it follows from the proof of Corollary 3.3 and the remarks immediately thereafter that $VC^+V = V$.

Corollary 3.7(a). $U^+ = G$ if and only if HV is idempotent and VH is hermitian.

Proof: The result follows from Corollary 3.7 by showing that HV idempotent and VH hermitian imply $C = V$, and conversely.

If $C = V$, then $H = V^+$, by Corollary 3.4, and we have $HV = V^+V$ idempotent and $VH = VV^+$ hermitian from Theorem 2.1.

Conversely, suppose that HV is idempotent and VH is hermitian. Now we know from the proof of Corollary 3.6(a) that HV idempotent implies $VC^+V = V$. Hence $H = C^+$, by Corollary 3.3, and so $VHV = V$,

$$HVH = C^+VC^+ = C^+CC^+ = C^+ = H,$$

and

$$(HV)^* = (C^+C)^* = C^+C = HV.$$

Then with $(VH)^* = VH$, by hypothesis, H and V satisfy the defining equations for the pseudoinverse. Therefore, $H = V^+$, but $H = C^+$. Hence $C = V$, by (P1).

For the special case $A = [A_{k-1}, a_k]$ with $A^+ = \begin{matrix} G_{k-1} \\ h_k \end{matrix}$ in which G_{k-1} has $k - 1$ rows and h_k is a single row, Corollaries 3.5 and 3.6(a) combine to give a form for A_{k-1}^+ corresponding to the representation for A^+ in (1) and (2). Since $h_k a_k$ is a scalar,we have

$$A_{k-1}^+ = \begin{cases} G_{k-1}[I + (1 - h_k a_k)^{-1} a_k h_k] , & \text{if } h_k a_k \neq 1 , \\ G_{k-1}(I - h_k^+ h_k) , & \text{if } h_k a_k = 1 . \end{cases}$$

A form for V^+ corresponding to each representation for U^+ follows at once from the dual symmetry noted above. In each case we simply inter-change U and V, G and H, and replace C by $\tilde{C}$.

3.3 Representations for the Pseudo Inverse of Sums of the Form $UU^* + VV^*$

The purpose of this section is to present representations for the pseudo inverse of certain sums of matrices. Consider matrices of the form $UU^* + VV^*$. Observe first that this sum is defined if and only if U and V have the same number of rows. The assumption that U and V have the same number of rows. The assumption that U and V have the same number of rows, is implicit throughout the following considerations.

Let A be any matrix with n columns partitioned as $A = (U, V)$, where U and V are submatrices with k and $n - k$ columns respectively, $0 \leq k \leq n$.

It is computational to confirm that A^+ can be written in the form

$$A^+ = \begin{bmatrix} U^+ - U^+VC^+ - U^+V\,(I - C^+C)\,KV^*U^{+*}U^+(I - VC^+) \\ C^+ + (I - C^+C)\,KV^*U^{+*}U^+\,(I - VC^+) \end{bmatrix} \qquad (1)$$

where

$$C = (I - UU^+)\, V \tag{2}$$

and

$$K = [I + (I - C^+C)\, V^*U^{+*}U^+V\, (I - C^+C)]^{-1} \tag{3}$$

A representation for $(UU^* + VV^*)$ is now obtained.

Theorem 3.5. For any matrices, U and V, the pseudoinverse of the sum $UU^* + VV^*$ can be written in the form

$$(UU^* + VV^*)^+ = (I - C^{+*}V^*)U^{+*}\,[I - U^+V\,(I - C^+C)\,KV^*U^{+*}]U^+$$
$$(I - VC^+) + C^{+*}C^+$$

where C and K are as defined in (2) and (3).

Proof: Let U and V be any matrices, and let $A = (U, V)$. Then $UU^* + VV^* = AA^*$, and it follows from (P11) that $(UU^* + VV^*)^+ = A^{+*}A^+$. The above representation for $(UU^* + VV^*)^+$ is obtained by using A^+ from (1) and block multiplication to form the product $A^{+*}A^+$. For this purpose let

$$M = I - U^+V\,(I - C^+C)\,KV^*\,U^{+*} \tag{4}$$

and

$$N = (I - C^+C)\,KV^*\,U^{+*}U^+\,(I - VC^+)\ ,$$

so that (1) can be written as

$$A^+ = \begin{bmatrix} MU^+\,(I - VC^+) \\ \\ C^+ + N \end{bmatrix} \tag{5}$$

Now by the defining equations in theorem 2.1, $I - C^+C$ is hermitian and $(I - C^+C)\ C^+ = 0$. Consequently, we have $N^*C^+ = 0$ and so

$$(C^+ + N)^* (C^+ + N) = C^{+*}C^+ + N^*N\ . \tag{6}$$

Next observe that since $I - C^+C$ is idempotent and commutes with the matrix K,

$$M^*M = I - 2U^+V (I - C^+C)\ KV^*U^{+*} + U^+V (I - C^+C)K (I - C^+C)\ V^*$$
$$U^{+*}\ U^+V (I - C^+C)\ KV^*U^{+*}$$

or

$$M^*M = I - U^+V (I - C^+C)\ KV^*U^{+*} - U^+V (I - C^+C)\ K^2V^*U^{+*}, \tag{7}$$

where

$$(I - C^+C)\ V^*U^{+*}U^+V (I - C^+C)\ K = I - K$$

by the definition of K. Finally, observing that multiplication of the last term in (7) on the left by $(I - C^{+*}V^*)U^{+*}$ and on the right by $U^+ (I - VC^+)$ gives $-N^*N$, then (5), (6), and (7) combine to give

$$A^{+*}A^+ = (I - C^{+*}V^*)\ U^{+*}MU^+ (I - VC^+) + C^{+*}C^+\ .$$

Replacing M by the expression in (4) yields the representation for $(UU^* + VV^*)^+$.

We now state and prove five special cases where the general form for A^+ given in theorem 3.5 simplifies.

Let

$$K_1 = (I + V^*U^{+*}U^+V)^{-1}\ .$$

Then we have

Corollary 3.8.

$$(UU^* + VV^*)^+ = (I - C^{+*}V^*)\ U^{+*}U^+\ (I - VC^+) - U^{+*}U^+V\ (I - C^+C)\ K_1V^*U^{+*}U^+ + C^{+*}C^+ \tag{8}$$

if and only if C^+C and $V^*U^{+*}U^+V$ commute.

Proof: If C^+C and $V^*U^{+*}U^+V$ commute, then

$$(I - C^+C)K = K_1\ (I - C^+C) = (I - C^+C)\ K_1\ (I - C^+C). \tag{9}$$

Therefore

$$U^+V\ (I - C^+C)\ KV^*U^{+*}U^+VC^+$$
$$= U^+V\ (I - C^+C)\ K_1V^*U^{+*}U^+V\ (I - C^+C)\ C^+ = 0$$

and dually

$$C^{+*}V^*U^{+*}U^+V\ (I - C^+C)\ KV^*U^{+*} = 0$$

in the representation for $(UU^* + VV^*)^+$, Theorem 3.5, which reduces to (8).

Conversely, suppose that $(UU^* + VV^*)^+$ has the form given in (8). Them combining the relations $C^+U = 0$, $C^+V = C^+C$, the definition of K_1 and the defining equations of Theorem 2.1 now gives

$$(UU^* + VV^*)^+\ (UU^* + VV^*) = (I - C^{+*}V^*)\ UU^+ + (I - C^{+*}V^*)\ U^{+*}U^+V\ (I - C^+C)V^* - U^{+*}U^+V\ (I - C^+C)\ K_1V^*UU^+ - U^{+*}U^+V\ (I - C^+C)\ (I - K_1)V^* + C^{+*}V^*$$

which reduces to

$$(UU^* + VV^*)^+ (UU^* + VV^*) = UU^+ + CC^+ - C^{+*}V^*U^{+*}U^+V\ (I - C^+C)(V^* + U^{+*}U^+V$$

$$(I - C^+C)\ K_1C^*$$

upon simplification, using the fact that $V^*UU^+ = V^* - C^*$. Continuing in the same manner yields

$$(UU^* + VV^*)\ (UU^* + VV^*)^+ (UU^* + VV^*) = UU^* + VV^*UU^+ + VV^*CC^+$$

$$- VV^*C^{+*}V^*U^{+*}U^+V\ (I - C^+C)V^* + UU^+V\ (I - C^+C)\ K_1C^* + VV^*U^{+*}U^+V$$

$$(I - C^+C)\ K_1C^*$$

or

$$(UU^* + VV^*)\ (UU^* + VV^*)^+ (UU^* + VV^*) = UU^* + VV^* - VC^+CV^*U^{+*}U^+V$$

$$(I - C^+C)V^* + V\ (I - C^+C)\ K_1C^* + VV^*U^{+*}U^+V\ (I - C^+C)\ K_1C^*\ ,$$

where again the definition of C is employed and we have used the relation

$$VV^*CC^+ = (CC^+VV^*)^* = (CV^*)^* = VC^* = VV^* - VV^*UU^+\ .$$

But

$$(UU^* + VV^*)\ (UU^* + VV^*)^+ (UU^* + VV^*) = UU^* + VV^*\ .$$

Therefore

$$- VC^+CV^*U^{+*}U^+V\ (I - C^+C)V^* + V\ (I - C^+C)\ K_1C^* + VV^*U^{+*}U^+V\ (- C^+C)$$

$$K_1C^* = 0\ . \tag{10}$$

Multiplying (10) on the left by C^+ and on the right by $U^{+*}U^+VC^+C$ now gives

$$C^+CV^*U^{+*}U^+V\ (I - C^+C)\ V^*U^{+*}U^+VC^+C = 0 \tag{11}$$

Taking $B = C^+CV^*U^{+*}U^+V\ (I - C^{+C})$ with C^+C hermitian and idempotent, (11) becomes $BB^* = 0$. Hence $B = 0$ and thus

$$C^+CV^*U^{+*}U^+V = C^+CV^*U^{+*}U^+VC^+C. \tag{12}$$

Consequently, with both C^+C and $V^*U^{+*}U^+V$ hermitian, the right hand side of (12) is hermitian and

$$C^+CV^*U^{+*}U^+V = (C^+CV^*U^{+*}U^+V)^* = V^*U^{+*}U^+VC^+C$$

as asserted.

<u>Corollary 3.9</u>. $(UU^* + VV^*)^+ = U^{+*}U^+ - U^{+*}U^+ - U^{+*}U^+VK_1V^*U^{+*}U^+ + C^{+*}C^+ \quad (13)$

if and only if $VC^+V = C$.

<u>Proof</u>: If $VC^+V = C$, then $VC^+ = VC^+CC^+ = VC^+VC^+ = CC^+$ and $U^+VC^+ = U^+CC^+ = 0$. Thus C^+C and $V^*U^{+*}U^+V$ commute, and (13) follows directly from (8).

Conversely, if (13) holds, then we can proceed as in the proof of Corollary 3.8 to form

$$(UU^* + VV^*)^+(UU^* + VV^*) = UU^+ + U^{+*}U^+VK_1C^* + C^{+*}V^*$$

and

$$(UU^* + VV^*)(UU^* + VV^*)^+(UU^* + VV^*) = UU^* + VV^*UU^+ + UU^+VK_1C^* + VV^*U^{+*}U^+VK_1D^* + VC^+VV^* .$$

$$(UU^* + VV^*)(UU^* + VV^*)^+ (UU^* + VV^*)$$

$$= UU^* + UU^+V (I - C^+C)V^* + V (I - C^+C)V^*UU^+$$

$$+ V (I - C^+C)V^*U^{+*}U^+V (I - C^+C)V^* + VC^+CV^* ,$$

from which it follows that

$$VV^* = 2V (I - C^+C)V^*$$

$$+ V (I - C^+C)V^*U^{+*}U^+V (I - C^+C)V^* + VC^+CV^* ,$$

or

$$0 = V (I - C^+C)V^* + V (I - C^+C)V^*U^{+*}U^+V (I - C^+C)V^*. \tag{16}$$

Since $I - C^+C$ is idempotent and $VV^*V^{+*} = V$ implies $CV^*V^{+*} = C$, multiplication of (16) on the right by $V^{+*}K$ gives

$$0 = V (I - C^+C) [I + (I - C^+C)V^*U^{+*}U^+V (I - C^+C)]K = V - VC^+V$$

and so $VC^+V = V$.

Corollary 3.11

$$(UU^* + VV^*)^+ = U^{+*}U^+ + V^{+*}V^+ \tag{17}$$

if and only if $C = V$.

Proof: If $C = V$, $VC^+V = VV^+V = V$ and $(UU^* + VV^*)^+$ can be written in the form given in Corollary 3.10. Also, $C = V$ implies $UU^+V = 0$. Hence $U^+V = 0$ and (15) reduces to (17).

Conversely, if (17) holds, then multiplying the relationship

$$(UU^* + VV^*)(U^{+*}U^+ + V^{+*}V^+)(UU^* + VV^*) = UU^* + VV^*$$

on the left by CC^+ and on the right by $C^{+*}C^*$, gives $CV^*U^{+*}U^+VC^* = 0$. Therefore $U^+VC^* = 0$ and

$$C = CC^+C = (I - UU^+)VC^+C = VC^+C - UU^+VC^*C^{+*} = VC^+C = VC^+V .$$

Applying Corollary 3.9 and the fact that the generalized inverse is unique, we have

$$V^{+*}V^+ = - U^{+*}U^+V\, K_1V^*U^{+*}U^+ + C^{+*}C^+$$

by equating the right hand sides of (13) and (17). Multiplication on the left by VV^* now gives

$$VV^+ = - V\,(I - K_1)V^*U^{+*}U^+ + VC^+ , \tag{18}$$

which provides the essential relation required to complete the proof.

Multiplying (18) on the right by V and using $VC^+V = C$ yields

$$V = C - V\,(I - K_1)V^*U^{+*}U^+V ,$$

or

$$UU^+V - V\,(I - K_1) + VV^*U^{+*}U^+V = 0 , \tag{19}$$

by the definition of C and K_1. Then we have

$$U^+V(K_1 + V^*U^{+*}U^+V) = 0 \tag{20}$$

by multiplying (19) on the left by U^+. Since K_1 positive definite and $V^*U^{+*}U^+V$ positive semidefinite imply that $(K_1 + V^*U^{+*}U^+V)^{-1}$ exists, it follows from (20) that $U^+V = 0$. Therefore $C = V$.

Corollary 3.12

$$(UU^* + VV^*)^+ = U^{+*}U^+ - U^{+*}U^+VK_1V^*U^{+*}U^+ . \tag{21}$$

if and only if $C = 0$.

Proof: If $C = 0$, then $C^+ = 0$, $VC^+V = C$, and (21) follows directly from the expression for $(UU^* + VV^*)^+$ in Corollary 3.9.

Conversely, if (21) holds, we have

$$(UU^* + VV^*)(UU^* + VV^*)^+ = UU^+ + CK_1V^*U^{+*}U^+ .$$

Since both the left hand side of this expression and UU^+ are hermitian, then

$$CK_1V^*U^{+*}U^+ = (CK_1V^*U^{+*}U^+)^* = U^{+*}U^+VK_1C^* .$$

Multiplying on the left by $I - UU^+$ and on the right by VC^*C^{+*} now gives

$$C(I - K_1)C^*C^{+*} = 0 ,$$

and so

$$C - CK_1C^*C^{+*} = 0 . \tag{22}$$

Forming $(UU^* + VV^*)(UU^* + VV^*)^+(UU^* + VV^*)$ and setting the resulting expression equal to $UU^* + VV^*$, it follows from

$$UU^* + UU^+VV^* + CK_1V^*UU^+ + C(I - K_1)V^* = UU^* + VV^*$$

that $CK_1C^* = 0$, which combines with (22) to give $C = 0$.

Numerical examples of matrices U and V for which $C^{+}C$ and $V^{*}U^{+*}U^{+}V$ do not commute and examples for which the conditions in Corollaries 3.8 to 3.12 hold are easily constructed. In fact, examples can be constructed using only matrices with elements zero or unity.

3.4 Pseudo Inverses of Sums of the Form $U + V$

Suppose now that U and V are matrices of the same size. Then we can consider representations for the generalized inverse of the sum $U + V$. For the special case of *-orthogonal matrices (that is, where U and V are matrices with both $UV^{*} = 0$ and $V^{*}U = 0$, we have shown in (P25) that $(U + V)^{+} = U^{+} + V^{+}$. In this section we develop representations for the generalized inverse of the sum $U + V$, where U and V are arbitrary rectangular matrices satisfying only the single condition $UV^{*} = 0$. Clearly, by applying the results to U^{*} and V^{*}, representations for $(U + V)^{+}$ when $U^{*}V = 0$ follow by symmetry. (P25) will again be established as a special case. (Corollary 3.16)

Consider any matrices U and V with $UV^{*} = 0$. Then

$$(U + V)(U + V)^{*} = UU^{*} + VV^{*} ,$$

and it follows from (P10) that

$$\begin{aligned}(U + V)^{+} &= (U + V)^{*} (UU^{*} + VV^{*})^{+} \\ &= U^{*}(UU^{*} + VV^{*})^{+} + V^{*} (UU^{*} + VV^{*})^{+} \qquad (1)\end{aligned}$$

Now from Theorem 3.5 we have a general form for $(UU^{*} + VV^{*})^{+}$ which can be substituted directly into (3.2). Alternatively, note that applying (P10) to the partitioned matrix $A = [U, V]$ gives

$$A^+ = A^*(AA^*)^+ = \begin{bmatrix} U^*(UU^* + VV^*)^+ \\ \\ V^*(UU^* + VV^*)^+ \end{bmatrix} \tag{2}$$

Since A^+ is unique, corresponding submatrices in (3.1) and (2) must be equal. Substitution of the expressions thus obtained for $U^*(UU^* + VV^*)^+$ and $V^*(UU^* + VV^*)^+$ into (3.54) gives

Theorem 3.6. For any matrices U and V such that $UV^* = 0$,

$$\begin{aligned}(U + V)^+ &= U^+ - U^+VC^+ - U^+V\,(I - C^+C)\,KV^*U^{+*}U^+\,(I - VC^+) + C^+ \\ &\quad + (I - C^+C)\,KV^*U^{+*}U^+\,(I - VC^+) \\ &= U^+ + (I - U^+V)\,[C^+ + (I - C^+C)\,KV^*U^{+*}U^+\,(I - VC^+)].\end{aligned}$$

The same five necessary and sufficient conditions employed in corollaries 3.8 - 3.12 are also applicable to establish special cases of the representation in Theorem 3.6. Since a proof of sufficiency in each of the following corollaries is obtained by taking the corresponding special representation for $A^+ = (U, V)^+$ developed in section II and forming the sum indicated in (1), only the necessity of each condition will be established. For this purpose we first note that with $UV^* = 0$, we have not only the relations $U^+C = 0$, $C^+U = 0$, and $C^+V = C^+C$, which hold for every U and V, but now also

$$UC^+ = UC^*C^{+*}C^+ = 0,$$

which implies

$$CU^+ = 0, \qquad CU^* = 0 \tag{3}$$

and $(I + V^*U^{+*}U^+V)\ U^* = U^*$, which implies

$$K_1U^* = U^*, \quad K_1U^+ = U^+ \tag{4}$$

Corollary 3.13. If $UV^* = 0$, then

$$\begin{aligned}(U + V)^+ = U^+ - U^+VC^+ - U^+V\ (I - C^+C)\ K_1V^*U^{+*}U^+ \\ + C^+ + (I - C^+C)\ K_1V^*U^{+*}U^+\end{aligned} \tag{5}$$

if and only if C^+C and $V^*U^{+*}U^+V$ commute.

Proof: (Necessity.) Suppose that $(U + V)^+$ has the form given in (5). Then it follows by equating this expression and the expression in Theorem 3.6 that

$$\begin{aligned}(I - U^+V)\ (I - C^+C)\ KV^*U^{+*}U^+\ (I - VC^+) \\ = (I - U^+V)(I - C^+C)\ K_1V^*U^{+*}U^+.\end{aligned}$$

Multiplication on the right by $V(I - V^*U^{+*})$ now gives

$$\begin{aligned}(I - U^+V)\ (I - C^+C)\ K\ (I - C^+C)\ (I - V^*U^{+*}) \\ = (I - U^+V)\ (I - C^+C)\ K_1\ (I - V^*U^{+*}),\end{aligned} \tag{6}$$

where we have used the fact that $I - C^+C$ commutes with K and the definitions of K and K_1. Since the left hand side of (6) is hermitian, then

$$\begin{aligned}(I - U^+V)\ (I - C^+C)\ K_1\ (I - V^*U^{+*}) \\ = (I - U^+V)\ K_1(I - C^+C)(I - V^*U^{+*}),\end{aligned}$$

and so

$$C^+CK_1\ (I - C^+C)\ (I - V^*U^{+*}) = 0, \tag{7}$$

where the second equation is obtained from the first by multiplication on the left side by C^+C and use of (3) and Theorem 2.1. Multiplying (7) on the right by U^* yields

$$C^+CK_1U^* - C^+CK_1\ (I - C^+C)V^*U^{+*}U^* = 0$$

or

$$C^+CK_1(I - C^+C)V^* = 0,$$

by (3), (4) and the definition of C. Hence $C^+CK_1\ (I - C^+C) = 0$, and it follows from the relation

$$C^+CK_1 = C^+CK_1C^+C$$

that C^+C and $V^*U^{+*}U^+V$ commute.

<u>Corollary 3.14</u>. If $UV^* = 0$, then

$$(U + V)^+ = U^+ - U^+VK_1V^*U^{+*}U^+ + C^+ + K_1V^*U^{+*}U^+ \tag{8}$$

if and only if $VC^+V = C$.

<u>Proof</u>: (Necessity.) If (8) holds, then it follows from the relation $(U + V)\ (U + V)^+\ (U + V) = U + V$ that

$$U + CK_1V^*U^{+*} + UU^+V + C\ (I - K_1) + VC^+V = U + V,$$

and so

$$CK_1V^*U^{+*} + C(I - K_1) + VC^+C = C.$$

Multiplication on the right by U^+V now gives

$$C(I - K_1) = 0.$$

Hence, $CV^*U^{+*}U^+V = 0$, which implies $U^+VC^* = 0$, and we conclude that $VC^+V = C$ as in the proof of necessity in Corollary 3.11.

Corollary 3.15. If $UV^* = 0$, then

$$(U + V)^+ = U^+ - U^+VC^+ + C^+ \tag{9}$$

if and only if $VC^+V = V$.

Proof: (Necessity.) If $(U + V)^+$ has the form given in (9), then if follows by equating this expression and the expression in Theorem 3.6 that

$$- U^+V(I - C^+C)KV^*U^{+*}U^+(I - VC^+)$$

$$+ (I - C^+C)KV^*U^{+*}U^+(I - VC^+) = 0.$$

Multiplication on the right by $VK^{-1}(I - V^*U^{+*})$ now gives

$$(I - U^+V)(I - C^+C)V^*U^{+*}U^+V(I - C^+C)(I - V^*U^{+*}) = 0,$$

which implies

$$(I - U^+V)(I - C^+C)V^*U^{+*} = 0 .$$

Using the relations $UV^* = 0$ and $UC^+ = 0$ it follows, therefore, that

$$UU^+V\,(I - C^+C)V^*U^{+*}U^+ = 0.$$

Hence $UU^+V\,(I - C^+C) = 0$, and so

$$V - VC^+V = V\,(I - C^+C)$$

$$= V\,(I - C^+C) - UU^+V\,(I - C^+C) = C\,(I - C^+C) = 0\,.$$

<u>Corollary 3.16</u>. If $UV^* = 0$, then

$$(U + V)^+ = U^+ + V^+$$

if and only if $C = V$.

<u>Proof</u>: (Necessity) Since $UV^* = 0$ implies that $UV^+ = 0$ and $VU^+ = 0$, then $(U + V)\,(U^+ + V^+)\,(U + V) = U + V$ gives

$$UU^+V + VV^+U = 0\,.$$

Multiplying by V^+V on the right we have

$$UU^+VV^+V = UU^+V = 0\,,$$

and so $C = V$.

<u>Corollary 3.17</u>. If $UV^* = 0$, then

$$(U + V)^+ = U^+ - U^+VK_1V^*U^{+*}U^+ + K_1V^*U^{+*}U^+ \tag{10}$$

if and only if $C = 0$.

<u>Proof</u>: (Necessity) If (10) holds, it follows from

$$(U + V)\ (U + V)^+\ (U + V) = U + V$$

that

$$U + UU^+V + CK_1V^*U^{+*} + C\ (I - K_1) = U + V\ ,$$

or

$$CK_1V^*U^{+*} + C\ (I - K_1) = C\ .$$

Multiplying by U^+V on the right we have $C\ (I - K_1) = 0$ as in the proof of Corollary 3.14. Hence $VC^+V = C$, and equating the expressions for $(U + V)^+$ in (8) and (10) gives $C^+ = 0$ and thus $C = 0$.

Observe in Corollary 3.16 that $C = V$ implies $U^*V = 0$, and conversely. When combined with the hypothesis of the corollary, we then have that $(U + V)^+ = U^+ + V^+$ if $UV^* = 0$ and $U^*V = 0$, that is, if U and V are *-orthogonal matrices. Also observe that when $C = V$ in the hermitian case, Corollary 3.11, UU^* and VV^* are *-orthogonal and (3.17) can be written in the alternative form

$$(UU^* + VV^*)^+ = (UU^*)^+ + (VV^*)^+\ .$$

Finally, it should be noted that although proofs of sufficiency in Corollaries 3.8 to 3.12 can be constructed directly by taking the corresponding special representation for $A^+ = [U, V]^+$ from section II and forming $A^{+*}A^+$, reduction of the resulting expression to obtain the given form for $(UU^* + VV^*)^+$ is required in Corollaries 3.8, 3.9, and 3.12. Hence, unlike the proofs in Corollaries 3.13 to 3.17

sufficiency in these corollaries to Theorem 3.5 is more easily established by direct reduction of the general form for $(UU^* + VV^*)^+$.

We now consider two applications of representations for $(U + V)^+$. We first establish relationships between V^+ and the pseudo inverse of $C = (I - UU^+)V$, where U and V are arbitrary matrices, and show as a special case that $VC^+V = C$ is a necessary and sufficient condition for V to have a particular decomposition into a sum of *-orthogonal matrices. We then consider the partitioned matrix $A = (U, V)$ and employ Corollary 3.15 to obtain a simple derivation of the form for A^+ in (3.1).

As noted above, each representation for $(U + V)^+$ with $UV^* = 0$ has a corresponding dual form with $U^*V = 0$. If F and G are any matrices of the same size with $F^*G = 0$, and if $\tilde{C} = (I - F^+F)G^*$ and

$$\tilde{K} = [I + I - \tilde{C}^+\tilde{C})GF^+F^{+*}G^* (I - \tilde{C}^+\tilde{C})]^{-1} ,$$

then it is easily shown that the dual form for the representation in Theorem 3.6 is

$$(F + G)^+ = F^+ - \tilde{C}^{+*}GF^+ - (I - \tilde{C}^{+*}G)F^+F^{+*}G^*\tilde{K} (I - \tilde{C}^+\tilde{C})GF^+$$
$$+ \tilde{C}^{+*} + (I - \tilde{C}^{+*}G)F^+F^{+*}G^*\tilde{K} (I - \tilde{C}^+\tilde{C}). \qquad (11)$$

This form for $(F + G)^+$ can be used to establish a general relationship between V^+ and C^+.

Let U and V be any matrices with the same number of rows, and consider the decomposition

$$V = (I - UU^+)V + UU^+V = C + UU^+V. \qquad (12)$$

Since $C^*U = 0$, then $F = C$ and $G = UU^+V$ satisfy $F^*G = 0$. In this case

$$\tilde{C} = (I - C^+C)V^*UU^+ = (I - C^+C)V^* ,$$

and, with $I - C^+C$ idempotent and hermitian, $(I - C^+C)\tilde{C} = \tilde{C}$ implies $\tilde{C}^+ = \tilde{C}^+ (I - C^+C)$, and so $\tilde{C}^+\tilde{C} = \tilde{C}^+V^*$. Substitution in (11) now gives

<u>Theorem 3.7.</u>

$$\begin{aligned} V^+ &= C^+ - \tilde{C}^{+*}UU^+VC^+ \\ &= (I - \tilde{C}^{+*}UU^+V)C^+C^{+*}V^*UU^+\tilde{K}\,(I - \tilde{C}^+V^*)UU^+VC^+ \\ &\quad + \tilde{C}^{+*} + (I - \tilde{C}^{+*}UU^+V)C^+C^{+*}V^*UU^+\tilde{K}\,(I - \tilde{C}^+V^*). \end{aligned}$$

For this representation we have

$$\tilde{K} = [I + (I - \tilde{C}^+V^*)UU^+VC^+C^{+*}V^*UU^+\,(I - \tilde{C}^+V^*)]^{-1}$$

Now $\tilde{K}$ can be replaced by

$$\tilde{K}_1 = [I + UU^+VC^+C^{+*}V^*UU^+]^{-1}$$

if and only if $\tilde{C}^+V^*$ and $UU^+VC^+C^{+*}V^*UU^+$ commute, and special forms for V^+ are easily obtained which correspond to Corollaries to Theorems 3.5 and 3.6. Analogous to Corollaries 3.14 to 3.17 necessary and sufficient conditions for V^+ to simplify can be stated in terms of UU^+V, $\tilde{C}$, and $\tilde{C}^+$. Alternatively, we can proceed as follows to obtain special cases in terms of V, C, and C^+.

If $C = V$, then $UU^+V = 0$, $\tilde{C} = (I - V^+V)V^*UU^+ = 0$, and the representation for V^+ reduces to C^+. On the other hand, if $C = 0$, then $\tilde{C} = V^* = V^*UU^+$ and the representation for V^+ reduces to $\tilde{C}^{+*} = (UU^+V)^+$. In each of these cases the converse follows immediately.

For the cases $VC^+V = C$ and $VC^+ = V$ we can proceed directly. In the first case, however, it is interesting to observe that $VC^+V = C$ is a necessary and sufficient condition that (12) is a decomposition into *-orthogonal matrices. We have

<u>Lemma 3.1</u>. C and UU^+V are *-orthogonal if and only if $VC^+V = C$.

<u>Proof</u>: Since $C^*U = 0$, we only need to show that $UU^+VC^* = 0$ implies $VC^+V = C$, and conversely. But his is obvious by noting that $UU^+VC^* = UU^+VC^+CC^*$ and that

$$C = CC^+C = VC^+C - UU^+VC^+C = VC^+V - UU^+VC^*C^{+*} .$$

Combining the dual form of Corollary 3.16 and Lemma 3.1 now gives

<u>Corollary 3.18</u>.

$$V^+ = C^+ + (UU^+V)^+$$

if and only if C and UU^+V are *-orthogonal.

Suppose that we are given matrices U and V such that (3.65) is a decomposition of V into *-orthogonal matrices. Then we know from Lemma 3.1 that $VC^+V = C$, and it follows by multiplying the expression for V^+ in Corollary 3.18 on the left and right by V that

$$UU^+V = V(UU^+V)^+V . \tag{13}$$

Now since both V^+V and VV^+ are hermitian, $UU^+VV^+V = UU^+V$ implies $V^+V(UU^+V)^+ = (UU^+V)^+$ and, with $VV^+C = VV^+VC^+V = VC^+V = C$, $VV^+UU^+V = UU^+V$ implies $(UU^+V)^)VV^+ = (UU^+V)^+$. Therefore, multiplying (13) on the left and right by V^+ gives $V^+UU^+VV^+ = V^+V(UU^+V)^+VV^+ = (UU^+V)^+$. Conversely, if the representation in Corollary 3.18 holds with $(UU^+V)^+ = V^+UU^+VV^+$, then $C^+VV^+ = C^+$. Thus $VV^+C = C$ and $VC^+V = VV^+ (I - UU^+)V = VV^+C = C$. This establishes

<u>Corollary 3.18(a)</u>.

$$C^+ = V^+ - V^+UU^+VV^+$$

if and only if C and UU^+V are *-orthogonal.

For the case $VC^+V = V$ we have

<u>Corollary 3.19</u>.

$$V^+ = C^+ - C^+C^{+*}V^*UU^+\tilde{K}_1UU^+VC^+ + C^+C^{+*}V^*UU^+\tilde{K}_1 \tag{14}$$

if and only if $VC^+V = V$.

Proof: Since V^+V is hermitian and $CV^+V = C$, then $V^+VC^+ = C^+$. Thus if $VC^+V = V$, then $V^+V = V^+VC^+V = C^+V = C^+C$, $\tilde{C} = (I - C^+C)V^* = 0$, $\tilde{K}$ reduces to $\tilde{K}_1$, and (14) follows directly from the representation in Theorem 3.7.

Conversely, if (14) holds, we have by multiplication on the left by UU^+V that

$$UU^+VV^+ = UU^+VC^+ - (I - \tilde{K}_1)UU^+VC^+ + (I - \tilde{K}_1),$$

by definition of $\tilde{K}_1$ or

$$UU^+VV^+ = \tilde{K}_1UU^+VC^+ + 1 - \tilde{K}_1 .$$

Multiplying on the right by V and rearranging terms now gives

$$\tilde{K}_1V - \tilde{K}_1UU^+VC^+V = C.$$

Since UU^+ is idempotent and commutes with $\tilde{K}_1$, and $U^+C = 0$, it follows that

$$\tilde{K}_1UU^+V(I - C^+V) = 0.$$

Thus $(V - C)(I - C^+V) = 0$ and so $V = VC^+V$.

Although we have employed the form for A^+ in (3.1) to build up the pseudo inverse representations for various sums of matrices, it is clear that each representation could have been established by direct verification of the defining equations in Theorem 2.1. In particular, having established the form for $(U + V)^+$ in Corollary 3.15, we can close the loop by giving a simple derivation of A^+ in (3.1).

Observe first that any matrices, U and V, with the same number of rows satisfy the relation

$$V = UU^+V(I - C^+C) + VC^+C.$$

Now setting $U = [U, UU^+V(I - C^+C)]$ and $V = [0, VC^+C]$ it follows for any matrix $A = [U, V]$ that

$$A = \underline{U} + \underline{V},$$

where $\underline{UV}^* = UU^+V(I - C^+C)C^+CV^* = 0$. Rewriting $\underline{U}$ as

$$\underline{U} = U[I, U^+V(I - C^+C)],$$

where the second factor in the product has full row rank, it can be shown that

$$\underline{U}^+ = [I, U^+V(I - C^+C)]^+U^+ .$$

Application of the fact that $A^+ = (A^*A)^+A^*$ to the first factor of this product now gives

$$\underline{U}^+ = \begin{bmatrix} \hat{K} \\ \\ (I - C^+C)V^*U^{+*}\hat{K} \end{bmatrix} U^+ \qquad (15)$$

with

$$\hat{K} = [I + U^+V(I - C^+C)V^*U^{+*}]^{-1} .$$

Then $\underline{UU}^+ = UU^+$,

$$\underline{C} = (I - \underline{UU}^+)\underline{V} = [0, (I - UU^+)VC^+C] = [0, C], \quad \underline{C}^+ = \begin{matrix} 0 \\ C^+ \end{matrix} ,$$

and so

$$\underline{VC^+V} = VC^+[0, VC^+C] = [0, VC^+C] = \underline{V} .$$

From Corollary 3.15 we have, therefore, that

$$A^+ = (\underline{U} + \underline{V})^+ = \underline{U} - \underline{U}^+\underline{VC}^+ + \underline{C}^+ . \qquad (16)$$

Finally, observing that $\hat{K}$ can be rewritten as

$$\hat{K} = I - U^{+}V\,(I - C^{+}C)KV^{*}U^{+*},$$

where

$$K = [I + (I - C^{+}C)V^{*}U^{+*}U^{+}V\,(I - C^{+}C)]^{-1},$$

and

$$(I - C^{+}C)V^{*}U^{+*}\hat{K} = (I - C^{+}C)KV^{*}U^{+*},$$

$\underline{U}^{+}$ in (15) becomes

$$\underline{U}^{+} = \begin{bmatrix} U^{+} - U^{+}V\,(I - C^{+}C)KV^{*}U^{+*}U^{+} \\ \\ (I - C^{+}C)KV^{*}U^{+*}U^{+} \end{bmatrix},$$

which combines with $\underline{C}^{+}$ and $\underline{VC}^{+} = VC^{+}$ to give the representation for A^{+} in (3.1) directly from (16).

We now give some concluding remarks on computational forms. If V is a single column, and either $C = 0$ or $C \neq 0$ and $C^{+}C = (C^{*}C)^{-1}C^{*}C = 1$. If we denote this special case by writing $V = \underline{v}$, $\underline{c} = (I - UU^{+})\underline{v}$ and $\underline{k}_1 = (1 + \underline{v}^{*}U^{+*}U^{+}\underline{v})^{-1}$, Corollaries 3.10 and 3.12 combine to give

$$(UU^{*} + \underline{vv}^{*})^{+} = \begin{cases} (I - \underline{c}^{+*}\underline{v}^{*})U^{+*}U^{+}\,(I - \underline{vc}^{+}) + \underline{c}^{+*}\underline{c}^{+}, & \text{if } \underline{c} \neq 0 \\ \\ U^{+*}U^{+} - \underline{k}_1 U^{+*}U^{+}\underline{vv}^{*}U^{+*}U^{+} & \text{if } \underline{c} = 0. \end{cases} \tag{17}$$

By the remarks after Corollary 3.17, concerning reduction of the general representation for $(UU^* + VV^*)^+$, it follows that forming the expression in (17) when $\underline{c} \neq 0$ is equivalent to applying formula (2.1), (2.2) to obtain $A^+ = [U,\underline{v}]^+$ and forming $A^{+*}A^+$. When $\underline{c} = 0$, however, application of the expression in (17) does not require direct formation of the submatrix $\underline{k}_1 U^+ \underline{vv}^* U^{+*} U^+$ employed in the formula for $A^+ = [U, \underline{v}]^+$, but only the formation of $\underline{k}_1 (\underline{v}^* U^{+*} U^+)^* (\underline{v}^* U^{+*} U^+)$.

CHAPTER 4

THE SCROGGS-ODELL PSEUDO INVERSE

4.1 Introduction

The definitions given in Chapter 2 fail to inherit an important spectral property of the inverse of a non-singular matrix. The property to which we refer is:

If a matrix A is nonsingular and if μ is an eigenvalue of A corresponding to the eigenvector x, then μ^{-1} is an eigenvalue of A^{-1} corresponding to x. Drazin [37] defined a generalized inverse of a matrix which preserves this spectral property of the inverse for the generalized inverse as far as it is possible to preserve it for a singular matrix. That is, if μ is a non-zero eigenvalue of the matrix A corresponding to the eigenvector x, then μ^{-1} is an eigenvalue of the pseudo inverse of A corresponding to the eigenvector x. Drazin defined the pseudo inverse A^D (D for Drazin) as follows: If J is the Jordan canonical form of a square matrix A, we have, of course $A = PJP^{-1}$. Now the Jordan matrix J can be regarded as the direct sum of a number of matrices J_i corresponding to the distinct eigenvalues λ_i of A. J_i is nonsingular if $\lambda_i \neq 0$ and nilpotent if $\lambda_i = 0$. Let J^D be the direct sum of $J^D{}_i$, where $J^D{}_i = J^{-1}$ if $\lambda_i \neq 0$ and $J^D{}_i$ is a null matrix if $\lambda_i = 0$. Finally, let $A^D = PJ^DP^{-1}$. Then it was shown that A^D was the unique matrix satisfying the three conditions:

$$A^{K+1}A^D = A^K \text{ for some positive integer } K.$$

$$AA^D = A^DA$$

$$A(A^D)^2 = A^D$$

Drazin shows that A^D is unique. However, it is not true in general that $(A^D)^D = A$, and it may occur that $A_1{}^D = A_2{}^D$ when $A_1 \neq A_2$. Also, $AA^DA = A$ only when A has generalized null vectors of height at most one, i. e. each J_i corresponding to $\lambda_i = 0$ is a null matrix in the above definition.

Recently, Odell and Scroggs [68] defined a pseudo inverse on which attention is focused in this chapter.

Throughout the discussion that follows it will be assumed that A is an n by n complex matrix representation of an operator on the n-dimensional Hilbert space X. The definition adopted here of a finite dimensional Hilbert space is that it is a finite dimensional, complete, complex inner product space.

A vector x is said to be a generalized eigenvector of A of height k, $k > 0$, corresponding to the eigenvalue u if and only if $(A - uI)^{k-1}x \neq 0$ and $(A - uI)^kx = 0$. A vector x is said to be a generalized eigenvector of maximal height corresponding to u if and only if there exists a positive integer k such that x is a generalized eigenvector of height k of A corresponding to u and $x \notin R(A - uI)$, the range of $A - uI$. If x is a generalized eigenvector of A corresponding to zero, we will say that x is a generalized null vector of A. If x is a generalized eigenvector of height k for A, then the sequence of vectors $(A - uI)^j x$, $j = 0, 1, 2, \ldots, k-1$, is said to be a chain of generalized eigenvectors of length k. In the definition given below for a pseudo inverse, use is made of the Jordan form of a matrix. Let C be the Jordan form of the matrix A, then there exists a matrix P such that $PAP^{-1} = C$. However, there are, in general, many choices for P. In order to insure uniqueness of the pseudo inverse, we shall place certain restrictions on P. The columns of P^{-1} are a basis for X. These columns are maximal chains of generalized eigenvectors of A. We restrict the possible choices for P by putting orthogonality restrictions on the columns of P^{-1}. The following restriction will be referred to as condition (0) for P with respect to A or simply as condition (0) when it is clear from the context what is meant.

Condition (0): Any generalized null vectors of maximal height, say k, of A which appear as columns of P^{-1} are mutually orthogonal and orthogonal to all generalized null vectors of A which are of height less than k.

It should be noted that if $PAP^{-1} = C$ where C is a Jordan form of A, then the fact that P satisfies condition (0) with respect to A implies the above orthogonality restrictions on the columns of P^{-1}.

We shall use the symbols $R(A)$ and $N(A)$ for the range and null space of A, respectively. Also, if U is a subspace of X its orthogonal complement will be designated by $U^{\perp}$.

4.2 Definition, Properties and an Application

We now define the Scroggs-Odell pseudoinverse.

Definition 4.1: Let A be an n by n matrix with Jordan canonical form C. Then there exists a non-singular matrix P satisfying condition (0) such that $PAP^{-1} = C$. Define C^I to be the matrix such that

$$C^I C = P_{R(C^I)} , \tag{1}$$

$$CC^I = P_{R(C)} , \tag{2}$$

where P_M is the orthogonal projection on M. Then the pseudo inverse of A, A^+, is defined by

$$A^+ = P^{-1}C^I P. \tag{3}$$

It follows easily from the definition that if A is non-singular then $A^+ = A^{-1}$.

Theorem 4.1: Let A be an n by n complex matrix representation of an operator. Then there always exist a matrix P satisfying condition (0) such that $PAP^{-1} = C$, where C is a Jordan form of A.

Proof: Consider the ranks of the iterates of A, $r(A)$, $r(A^2)$, ..., $r(A^k)$. Let k be the smallest integer such that $r(A^k) = r(A^{k+1})$. Find a basis for $N(A)$. Compute a basis for $N(A^P) - N(A^{P-1})$, $P = 2, 3, \ldots, k-1$. From considerations of rank, we see that it is possible to find an orthonormal basis for $N(A^k) - N(A^{k-1})$ which is orthogonal to $N(A^{k-1})$ and, consequently, orthogonal to $N(A^P)$, $P = 1, 2, \ldots, k-2$. These basis vectors are generalized null vectors of maximal height k of A which are mutually orthogonal and orthogonal to all generalized null vectors of A of less height. If for any interger $m < k$, we have $r(A^{m-1}) - r(A^m) > r(A^m) - r(A^{m+1})$, then there are generalized null vectors of maximal height m. Now if x is a generalized null vector of maximal height greater than m, then some iterate of A operating on x belongs to $N(A^m)$. If $x_1, x_2, \ldots, x_q$ are those vectors in $N(A^m)$ which are images of vectors of maximal height greater than m, we complete this set to a basis by using mutually orthogonal vectors which are orthogonal to $x_1, x_2, \ldots, x_q$ and to $N(A^{m-1})$. Again an appeal to the rank of A^m shows that this is possible. We now have a basis for the generalized null space of A consisting of maximal chains of generalized null vectors of A. Since X is the direct sum of $N(A^k)$ and the generalized range of A, we can construct a canonical basis for the generalized range of A in the usual manner so that the representation of A in the above basis for X is a Jordan form for A.

We now state and prove three theorems which are interesting in their own right, but also needed later to establish that the above definition gives a unique pseudo inverse for the matrix A.

Theorem 4.2: If C is the Jordan canonical form of A, then

$$CC^IC = C \tag{4}$$

$$C^ICC^I = C^I \tag{5}$$

$$(C^IC)^* = C^IC \tag{6}$$

$$(CC^I)^* = CC^I . \tag{7}$$

Proof: Equations (1) and (2) define a pseudo inverse for C. This definition is the same as the definition of E. H. Moore [66]. Equations (4), (5), (6) and (7) are used by Penrose [70] to define a pseudo inverse. These have been shown to be equivalent in [6].

Theorem 4.3: If A is the n by n matrix representation of an operator on the n-dimensional Hilbert space X, then $X = R(C^I) \oplus N(C)$ and $R(C) \oplus N(C^I)$.

Proof: This follows directly from the work of Desoer and Whalen [36].

Theorem 4.4: C^IC and CC^I are diagonal matrices with diagonal elements either 1 or 0. Considering the $\{e_i\}_{i=1}^n$ basis for X, if $Ce_{i_h} = 0$, then the i_h-th diagonal element of C^IC is 0, if $Ce_{i_k} \neq 0$, then the i_k-th diagonal element of C^IC is 1. Similarly, if $C^Ie_{i_h} = 0$, then the i_h-th diagonal element of C^IC is zero, and if $C^Ie_{i_k} \neq 0$, then the i_k-th diagonal element of CC^I is 1.

Proof: Since $R(C^I) \oplus N(C) = X$, we can divide the basis into two disjoint sets $E' = \{e_{i_k}\}_{k=1}^{p}$ and $E'' = \{e_{i_h}\}_{h=1}^{r}$ with E' spanning $R(C^I)$ and E'' spanning $N(C)$. Now if $e_{i_h} \varepsilon N(C)$, then $C^I Ce_{i_h} = 0$. Therefore, every element in the i_h-th column of $C^I C$ must be zero. However, if $e_{i_k} \varepsilon R(C^I)$, i.e., $Ce_{i_k} \neq 0$, then by (1) $C^I Ce_{i_k} = e_{i_k}$. Hence, every element in the i_k-th column of $C^I C$ is zero except for the diagonal element and the i_k-th diagonal element must be 1. Hence $C^I C$ has the form described in the theorem. The verification of the form of CC^I is easily confirmed in a like manner.

Lemma 4.1: C^+ is the unique pseudo inverse of C.

Proof: Let C_1 be any Jordan form of C. We must show that if $C = P_1^{-1} C_1 P_1$ where P_1 satisfies condition (0) with respect to C, then $C^I = P_1^{-1} C_1^I P_1 = C^+$, where C^I and C_1^I are defined by (1) and (2). Clearly there exists a permutation matrix Q such that $C_1 = Q' CQ$. Therefore, $C = P_1^{-1} Q' CQP_1 = P^{-1} CP$, where $P = QP_1$. As P^{-1} represents a mere rearrangement of the columns of P_1^{-1}, condition (0) for P_1 with respect to C implies condition (0) for P with respect to C.

Thus, it suffices to show that if $C = P^{-1} CP$, where P satisfies condition (0) with respect to C, then $C^I = P^{-1} C^I P$. In view of the preceding paragraph, we may assume that C can be partitioned in the following manner

$$C = \begin{bmatrix} C_2 & 0 \\ 0 & C_3 \end{bmatrix}, \tag{8}$$

where C_2 is non-singular and C_3 consists of all of the Jordan blocks of C corresponding to the eigenvalue zero. Partitioning P^{-1}, we have

$$\begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \begin{bmatrix} C_2 & 0 \\ 0 & C_3 \end{bmatrix}^k = \begin{bmatrix} C_2 & 0 \\ 0 & C_3 \end{bmatrix}^k \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \tag{9}$$

for any positive integer k. For sufficiently large k,

$$\begin{bmatrix} C_2 & 0 \\ 0 & C_3 \end{bmatrix}^k = \begin{bmatrix} C_2{}^k & 0 \\ 0 & 0 \end{bmatrix} \tag{10}$$

The non-singularity of C_2 and equation (9) imply that $P_2 = 0$, $P_3 = 0$. Equation (9) for $k = 1$ is equivalent to $P_1C_2 = C_2P_1$ and $P_4C_3 = C_3P_4$. As a consequence of the non-singularity of P, both P_1 and P_4 are non-singular. Also, P_1 commutes with C_2 if and only if it commutes with $C_2{}^{-1}$.

Thus, in order to show that $C^I = P^{-1}C^IP$, it suffices to show that $C_3P_4 = P_4C_3$ implies that $P_4C'_3 = C'_3P_4$.

Let the Jordan blocks of C_3 be $C_{31}, C_{32}, \ldots, C_{3p}$, i.e., $C_3 = \text{diag}\,(C_{31}, C_{32}, \ldots, C_{3p})$ where C_{3i} is an n_i by n_i matrix all of whose elements are either zero or one. If $n_i > 1$, then the only non-zero elements are the elements of the diagonal above the principal diagonal. Partition P_4 in a manner conformal to the partitioning of C_3, i.e., $P_4 = (P_{ij})$, where P_{ij} is an n_i by n_j matrix. We assume that $P_4C_3 = C_3P_4$. Let $Q = P_4C_3$ and $R = C_3P_4$. Then the above partitioning of P_4 and C_3 produces, in a natural way, a partitioning of Q and R. If $Q - (Q_{ij})$ and $R = (R_{ij})$, then $(Q_{ij}) = (P_{ij}C_{3j})$ and $(R_{ij}) = (C_{3i}P_{ij})$. Now, from the nature of C_{3j}, we see that the first column of Q_{ij} is zero. The k-th column of Q_{ij} is the same as the (k-1)-th column of P_{ij} for $k = 2, 3, \ldots, n_j$. From the nature of C_{3i}, it follows that the last row of R_{ij} is zero and for $k = 1, 2, \ldots, n_i-1$, the k-th row of R_{ij} is the same as the (k+1)-th row of P_{ij}. By assumption, $(Q_{ij}) = (R_{ij})$. Thus, (1) every element of the first column of P_{ij} is zero except, possibly, the (1,1) element, (2) every element of the last row of P_{ij} is zero except, possibly, the (n_i, n_j) element, and (3) the elements of any given diagonal sloping downward to the right are equal.

Now consider the canonical basis, $\{e_i\}_{i=1}^{n}$. Suppose C_2 is a t by t matrix. The columns of P^{-1} after the t-th column form chains of generalized null vectors of C corresponding to the Jordan blocks of C_3. The column to the right in each chain is of maximal height. We can establish a one-to-one correspondence between the elementary vectors e_i for $i > t$ and the columns of

P^{-1} after the t-th column by making the i-th column of P^{-1} correspond to e_i. It is easily verified by actual multiplication that if e_i corresponds to a column of P^{-1} which is the l-th column of its chain, counting from the left, then e_i is a generalized null vector of C of height 1.

Let p be the right-hand column of the chain corresponding to the j-th Jordan block of C_3. Then by condition (0) p is orthogonal to all generalized null vectors of C of height less than n_j. These include all elementary vectors e_i corresponding to columns of P^{-1} whose ordinal position in their chain is less than n_j. It follows that the elements of p in the corresponding row position are zero. In view of (3), we can conclude that $P_{ij} = 0$ if $n_i < n_j$ and P_{ij} is a diagonal matrix if $n_i \geq n_j$. It remains to be shown that the diagonal elements are zero if $n_i > n_j$.

p has nonzero elements at most only in those row positions corresponding to columns of P^{-1} which are generalized null vectors of C of height n_j. Let there be m columns of height n_j belonging to chains of length greater than n_j. The submatrix consisting of these m columns is of rank m, since P^{-1} is nonsingular. Moreover, the nonzero elements in these columns are confined to those row positions corresponding to the ordinal positions in P^{-1} of the columns themselves. Deleting the zero rows would give a nonsingular m by m submatrix. The orthogonality of p to each of the m columns, as required by condition (0) implies that a linear combination of the rows of the latter submatrix vanishes. Therefore the coefficients in the linear combination vanish. But these include all of those elements of p which are the (n_j, n_j) elements of blocks P_{ij} for which $n_i > n_j$. In view of (3), the desired conclusion follows. Thus all the elements of P_4 are zero except those of the principal diagonal of square submatrices P_{ij}. Furthermore, for a given P_{ij}, the elements of the principal diagonal are equal. But if P_4 is of this form, then so is its transpose, $(P_4)'$. Also, any matrix of this form commutes with C_3. Thus $(P_4)'C_3 = C_3(P_4)'$ or, taking transposes, $P_4C'_3 = C'_3P_4$. Hence, $C^I = P^{-1}C^IP$. That is, $C^I = C^+$ and thus is the pseudo inverse of C according to the above definition. The uniqueness follows since C^I was shown to be unique [70].

In view of Lemma 4.1, in the sequel C^I will be used to designate the pseudo inverse of a matrix C in Jordan form, whether it is defined by equations (1) and (2), or by equation (3). We now establish that the pseudo inverse given by equation (3) of an arbitrary n by n matrix A exists and is unique.

Theorem 4.5: If A is an n by n matrix, then A^+ exists and is unique. Furthermore, A^+ satisfies the following

$$AA^+A = A \tag{11}$$

$$A^+AA^+ = A^+ , \tag{12}$$

Proof: The existence of C^I satisfying equations (1) and (2) is guaranteed by Theorem 4.2 and the work of Penrose [70]. The existence of A^+ then follows immediately from Theorem 4.1 and equation (3). Suppose $A = P^{-1}CP = P_2^{-1}C_1P_2$, where both P and P_2 satisfy condition (0). Then $C = PP_2^{-1}C_1P_2P^{-1}$. We wish to show that if P and P_2 each satisfy condition (0) with respect to A, then P_2P^{-1} satisfies condition (0) with respect to C. There exists a permutation matrix R such that $C_1 = R'CR$, and therefore

$$C = PP_1^{-1}CP_1P^{-1} ,$$

where $P_1 = RP_2$. As $P_2^{-1} = P_1^{-1}R'$ represents a mere rearrangement of the columns of P_2^{-1}, P_1 satisfies condition (0) if P_2 does. It is sufficient, therefore, to show that P_1P^{-1} satisfies condition (0) with respect to C if P and P_1 satisfy condition (0) with respect to A. We use p_i to designate row i of P and p^j to designate column j of P^{-1}. Similarly ${}_1p_i$ will designate row i of P_1 and ${}_1p^j$ will designate column j of P_1^{-1}. Suppose e_k is a generalized null vector of maximal height m of C. Then ${}_1p^k$ and p^k will be generalized null vectors of maximal height m of A. If $PP_1^{-1} = Q = (q_{ij})$, to confirm condition (0) for P_1P^{-1}, we need to show that if e_t, $t \neq k$ is a generalized null vector of maximal height m of C, then $\sum_{i=1}^{n} q_{it}\bar{q}_{ij} = 0$, and secondly if e_t, $t \neq k$ is a generalized null vector of C of height at most m, not of maximal height m, then $\sum_{i=1}^{n} q_{it}\bar{q}_{ik} = 0$. Now $q_{ij} = ({}_1p^j, p_i^*)$. We now prove the first part. By condition (0) for P_1, $({}_1p^t, {}_1p^k) = 0$. Let $\{p^{i_s}\}_{s=1}^{u}$ be the set of generalized null vectors of maximal height m which appear as columns of P^{-1}. Since $p = P^{-1}(Pp) = \sum_{i=1}^{m} (P,P_i^*)P^i$, we have

$$ {}_1p^t = \sum_{i=1}^{n} ({}_1p^t,\ p_i^*)\ p^i \tag{13}$$

$$ {}_1p^k = \sum_{i=1}^{n} \ {}_1p^k,\ p_i^*)\ p^i \tag{14}$$

where $({}_1p^t, p_i^*) = ({}_1p^k, p_i^*) = 0$ if $i \neq i_1, i_2, \ldots, i_u$, since by condition (0) ${}_1p_t$ and ${}_1p_k$ can be expressed uniquely as linear combinations of p^{i_s}, $s = 1, 2, \ldots, u$.

Hence,

$$0 = \Big(\sum_{i=1}^{n} ({}_1p^t,\ p_i^*)\ p^i,\ \ \sum_{i=1}^{n} ({}_1p^k,\ p_i^*)\ p^i\Big)$$

$$= \sum_{i=1}^{n} ({}_1p^t,\ p_i^*)\ \overline{({}_1p^k,\ p_i^*)}$$

$$= \sum_{i=1}^{n} q_{it}\ \bar{q}_{ik}$$

Therefore, $\sum_{i=1}^{n} q_{it}\, q_{ik} = 0$ as was to be shown. We now prove the second part. If e_t, $t \neq k$ is a generalized null vector of height at most m, then $({}_1p^t, {}_1p^k) = 0$ by condition (0). Now $({}_1p^t, p_i^*) = 0$ for $i = i_1, i_2, \ldots, i_u$, and $({}_1p^k, p_j^*) = 0$ if p^j is not a generalized null vector of maximal height m.

Hence,

$$0 = \sum_{j=1}^{n} ({}_1p^t,\ p_j^*)\ \overline{({}_1p^k,\ p_j^*)}$$

$$= \sum_{i=1}^{n} q_{it}\ \bar{q}_{ik}.$$

Therefore,

$$\sum_{i=1}^{n} q_{it}\ \bar{q}_{ik} = 0$$

as was to be shown.

Using Lemma 4.1, $C^I = PP_2^{-1}C_1^IP_2P^{-1}$. Thus, $P^{-1}C^IP = P_2^{-1}C_1^IP_2 = A^+$ and A^+ is unique.

Equations (11) and (12) follow from the definition and equations (4) and (5).

The question that naturally comes up now is whether a unique pseudo inverse could have been obtained with a less severe restriction on P than that given by condition (0). Suppose we only required that any generalized null vectors of maximal height k which appeared as columns of P^{-1} be orthogonal to all generalized null vectors of A which are of height less than k. The following is then a counterexample to Lemma 4.1, and thus to Theorem 4.5.

Let

$$C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The above condition is satisfied as $(0, 0, 1)'$ is the only column of P^{-1} of height two and it is orthogonal to all null vectors of C. It is easily verified by direct multiplication that $C = P^{-1}CP$. From Theorem 4 it follows that

$$C^I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

However, calculating $C^+ = P^{-1}C^IP$ we get

$$C^{+} = P^{-1}C^{I}P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 0 \end{bmatrix} \neq C^{I} .$$

Lemma 4.2: Let $PAP^{-1} = C$ where C is the Jordan canonical form for A. If x is a generalized eigenvector of height k for A corresponding to the eigenvalue u, then Px is a generalized eigenvector of height k for C corresponding to u.

Proof: We first note that

$$(A - uI)^{n} = (P^{-1}CP - uI)^{n} = [P^{-1}(C - uI)P]^{n} =$$

$$P^{-1}(C - uI)^{n}P$$

for any positive integer n. By assumption $(A - uI)^{k-1}x \neq 0$. This implies that $P^{-1}(C - uI)^{k-1}Px \neq 0$ and hence that $(C - uI)^{k-1}Px \neq 0$. Likewise $(A - uI)^{k}x = 0$ implies that $(C - uI)^{k}Px = 0$. Hence Px is a generalized eigenvector of height k for C corresponding to the eigenvalue u.

Theorem 4.6: If u is a non-zero eigenvalue of A and x is the corresponding eigenvector, then u^{-1} is a non-zero eigenvalue of A^{+} and x is the corresponding eigenvector. If A has rank r, then A^{+} has rank r.

Proof: If u is a non-zero eigenvalue of A with eigenvector x, then $Ax = ux$. Hence $A^{+}Ax = uA^{+}x$. But this implies that $P^{-1}C^{I}PP^{-1}CPx = uA^{+}x$, or $P^{-1}C^{I}CPx = uA^{+}x$. But if x is an eigenvector of A, then $y = Px$ is an eigenvector of C

corresponding to u by Lemma 4.2. From the form of $C^I C$ given by Theorem 4.4, it follows that $C^I Cy = y$. Thus

$$P^{-1}C^I CPx = x = uA^+x.$$

Since $u \neq 0$, the result then follows from the division by u. It follows from Theorem 4.3 that the rank of A is the same as the rank of A^+.

In general, the property that $(A^{-1})^{-1} = A$ does not carry over to the pseudo inverse defined above. The next theorem gives us necessary and sufficient conditions for $(A^+)^+$ to be equal to A provided that no two maximal chains of generalized null vectors of A are of the same length k for $k > 1$.

Theorem 4.7: Assume A is such that the length, k, of each cahin of generalized null vectors of A is different for $k > 1$. Then $(A^+)^+ = A$ if and only if there exist a matrix P such that $PAP^{-1} = C$, where C is a Jordan canonical form of A, and in addition to property (0), P has the property that for each chain of generalized null vectors of length, say k, of A appearing as columns of P^{-1}, the null vector of the chain is orthogonal to all the other generalized null vectors of height at most k which appear as columns of P^{-1}.

Proof: First we show that $(C^I)^+ = C$. Let $C_1, C_2, \ldots, C_{k-1}$ be the Jordan blocks of C corresponding to non-zero eigenvalues of A, and C_k be the matrix which is the direct sum of all Jordan blocks corresponding to the zero eigenvalue. Then

$$C = \text{diag}\,(C_1, C_2, \ldots, C_{k-1}, C_k)\ .$$

It follows from Theorem 4.4 that

$$C^I = \text{diag}\,(C_1^{-1}, C_2^{-1}, \ldots, C_{k-1}^{-1}, C_k^T)\ .$$

Let D be the Jordan canonical form of C^I with

$$D = \text{diag}\,(D_1, D_2, \ldots, D_{k-1}, D_k)$$

The summands of D can be chosen so that there exist P_i, $i = 1, 2, \ldots, k$ such that $C_i^{-1} = P_i^{-1}D_iP_i$, $i = 1, 2, \ldots, k-1$ and $C_k^T = P_k^{-1}D_kP_k$. Hence

$$(C^I)^+ = \text{diag}\,(P_1^{-1}D_1^{-1}P_1, P_2^{-1}D_2^{-1}P_2, \ldots, P_{k-1}^{-1}D_{i-1}^{-1}P_{k-1}, P_k^{-1}D_k^TP_k)$$

$$= \text{diag}\,(C_1, C_2, \ldots, C_{k-1}, C_k)$$

$$= C\ .$$

Hence $(C^I)^+ = C$.

Suppose P has the property described in the theorem. Then if $x_1, \ldots, x_k$ with $x_i = A^{i-1}x_1$ is a maximal chain of generalized null vectors of A which appear as columns of P^{-1}, then $x_k, \ldots, x_1$ with $x_{k-1} = (A^+)^i x_k$ is a maximal chain of generalized null vectors of A^+.

Partition C with $C = \text{diag}\,(C_1, \ldots, C_k)$ where the C_i, $i = 1, 2, \ldots, k$ are the Jordan blocks of C. If C_i is a Jordan block corresponding to a non-zero eigenvalue of A, let

Q_i be that matrix such that $C_i^{-1} = Q_i^{-1} D_i Q_i$, where D_i is the Jordan form of C_i^{-1}. If C_i is an n_i by n_i Jordan block corresponding to the eigenvalue zero, then let Q_i be the n_i by n_i matrix all of whose elements are zero except the elements on the diagonal from the $(1, n_i)$ position to the $(n_i, 1)$ position. Each of the elements on this diagonal is one. That is, in this case Q_i is a permutation matrix which reverses the order of the columns of a matrix when the matrix is multiplied on the right by Q_i. Also, in this case $Q_i^{-1} = Q_i$. Let $Q = \text{diag}\ (Q_1, Q_2, \ldots, Q_k)$. Now $C^I = Q^{-1}DQ$ where $D = \text{diag}\ (D_1, D_2, \ldots, D_k)$. Thus $A^+ = P^{-1}Q^{-1}DQP$. Since P has the property described in the theorem, QP has property (0) with respect to A^+.

Hence $(A^+)^+ = P^{-1}Q^{-1}D^IQP$. Also, Q has property (0) with respect to C^I. Consequently, $C = (C^I)^+ = Q^{-1}D^IQ$. Thus $A = P^{-1}CP = P^{-1}Q^{-1}D^IQP = (A^+)^+$.

Now assume that $(A^+)^+ = A$. Let D be a Jordan form for A^+. Then $P^{-1}C^IP = R^{-1}DR$, or

$$C^I = PR^{-1}DRP^{-1} \tag{15}$$

Also since $(A^+)^+ = A$, $A = R^{-1}D^IR = P^{-1}CP$, or

$$C = PR^{-1}D^IRP^{-1} \tag{16}$$

Let $S = RP^{-1}$. Partitioning C^I, D and S^{-1} and using the same sort of reasoning as in Lemma 1 we can replace equation (15) by

$$\begin{bmatrix} C_1^{\ I} & 0 \\ 0 & C_2^{\ I} \end{bmatrix} \begin{bmatrix} S_1^{\ -1} & 0 \\ 0 & S_2^{\ -1} \end{bmatrix} = \begin{bmatrix} S_1^{\ -1} & 0 \\ 0 & S_2^{\ -1} \end{bmatrix} \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \tag{17}$$

where $C_2^{\ I}$ and D_2 are the nilpotent parts of C^I and D, respectively. Now $C_1^{\ I} = C_1^{\ -1}$ and $C_2^{\ I} = C_2^{\ T}$. Similarly, equation (16) can be replaced by

$$\begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix} \begin{bmatrix} S_1^{\ -1} & 0 \\ 0 & S_2^{\ -1} \end{bmatrix} = \begin{bmatrix} S_1^{\ -1} & 0 \\ 0 & S_2^{\ -1} \end{bmatrix} \begin{bmatrix} D_1^{\ I} & 0 \\ 0 & D_2^{\ I} \end{bmatrix} \tag{18}$$

Without any loss of generality, we can assume that $C_2 = D_2$. A consequence of (17) is that

$$C_2^{\ T} S_2^{\ -1} = S_2^{\ -1} C_2 \, . \tag{19}$$

Also, from (18) we get that

$$C_2 S_2^{\ -1} = S_2^{\ -1} C_2^{\ T} \, . \tag{20}$$

Let $S_{ij} = (S_{kl})$ be a n_{ij} by m_{ij} partition block of $S_2^{\ -1}$. Suppose $n_{ij} \geq m_{im}$. It is easily verified by direct multiplication that equation (19) implies that every element of S_{ij} below the diagonal containing $s_{1,m_{ij}}$, $s_{2,m_{ij}-1}$, etc., is zero. Similarly, (20) implies that every element of S_{ij} above the diagonal containing $s_{n_{ij},1}$, $s_{n_{ij}-1,2}$, etc., is zero. If $n_{ij} \leq m_{ij}$, then (19) implies that every element of S_{ij} below the diagonal containing $s_{n_{ij},1}$, $s_{n_{ij}-1,2}$, etc., is zero, and (20) implies that every element of S_{ij} above the diagonal containing $s_{1,m_{ij}}$, $s_{2,m_{ij}-1}$, etc., is zero. Thus, the two equations, (19) and (20), imply that the partition blocks of S_2^{-1} are sero if the blocks are not square and are of a diagonal form otherwise. Since the chains are each of a different length, the only square blocks of S_2^{-1} are on the diagonal of S_2^{-1}.

Now let $A^j x_1 = s_{j+1}$, $j = 0, 1, \ldots, k-1$, be a maximal chain of generalized null vectors appearing as successive columns of P^{-1}. Since $D_2 = C_2$, there is a corresponding maximal chain of generalized null vectors of A^+, say $(A^+)^j y_1 = y_{j+1}$, $j = 0, 1, \ldots, k-1$, which are columns of R^{-1}. The fact that $(A^+)^+ = A$ implies that

$$A^{k-j} y_k = y_j, \quad j = 1, 2, \ldots, k$$

is a maximal chain for A. Thus y_1 is a null vector for A. Furthermore, cince R has property (0), y_1 is orthogonal to all generalized null vectors of A^+ of height at most k. From the form of S_2^{-1}, using the relationship $R^{-1} = P^{-1}S^{-1}$, one can deduce that y_k is a scalar multiple of the generalized null vector x_1 of maximal height k appearing as a column of P^{-1}. But, according to property (0) for P, x_1 is orthogonal to all other generalized null vectors of A of height at most k, (except itself). Let P_1^{-1} be the matrix obtained by replacing each maximal chain, $x_1, x_2, \ldots, x_k$ by the corresponding $y_k, y_{k-1}, \ldots, y_1$, respectively. Then P_1 has the property described in the theorem.

Without requiring that the length of each chain of generalized null vectors of length greater than one be different, one is able to establish only sufficient conditions for $(A^+)^+ = A$. The following is easily established from Theorem 4.7.

Corollary 4.1. If there exist a matrix P such that $PAP^{-1} = C$ where C is a Jordan form of A, and in addition to property (0), P has the property that for each chain of generalized null vectors of length, say k, of A appearing as columns of P^{-1}, the null vector of the chain is orthogonal to all the other generalized null vectors of height at most k which appear as columns of P^{-1}, then $(A^+)^+ = A$.

Lemma 4.3: Let $PAP^{-1} = C$ where C is the Jordan canonical form of A, and P satisfies condition (0). If $B = QAQ^*$ with $Q^* = Q^{-1}$, then PQ^* satisfies condition (0) with respect to B.

Proof: First we show that if x_i is a generalized null vector of maximal height of A which appears as a column of P^{-1}, then Qx_i is a generalized null vector of maximal height for B. Certainly $A^{k-1}x_i \neq 0$ implies that $Q^*B^{k-1}Qx_i \neq 0$. But, this implies that $B^{k-1}Qx_i \neq 0$. Also, $A^k x_i = 0$ implies that $Q^*B^kQx_i = 0$ which implies that $B^kQx_i = 0$. If $Qx_i \in R(B)$, there exist a vector y such that $By = Qx_i$. Hence $QAQ^*y = Qx_i$. Multiplying on the left by Q^* we have that $AQ^*y = x_i$ which implies that $x_i \in R(A)$. But $x_i \notin R(A)$ by assumption therefore, Qx_i is a generalized null vector of maximal height for B.

Partition P^{-1} into its columns, say, $P^{-1} = (x_1, x_2, \ldots, x_n)$. Then $QP^{-1} = (Qx_1, Qx_2, \ldots, Qx_n)$. Let $y_i = Qx_i$ be a column of QP^{-1}

which is a generalized null vector of maximal height k of B, and $y_i = Qx_j$ be a distinct column of QP^{-1} which is a generalized null vector of B of height at most k. Then $(y_i, y_j) = (Qx_i, Qx_j) = (x_i, x_j) = 0$. Hence PQ^* satisfies condition (0).

<u>Theorem 4.8</u>: If $PAP^{-1} = C$ and $B = QAQ^*$, then $B^+ = QA^+Q^*$.

<u>Proof</u>: $PAP^{-1} = C$ and $A = Q^*BQ$ implies that $PQ^*BQP^{-1} = C$. By Lemma 3, PQ^* satisfies condition (0). Hence, $PQ^*B^+QP^{-1} = C^I$, so that $PQ^*B^+QP^{-1} = PA^+P^{-1}$. This implies that $B^+ = QA^+Q^*$. But $PA^+P^{-1} = C^I$.

We now impose some additional restrictions on the columns of P^{-1} where $PAP^{-1} = C$ and C is the Jordan canonical form of A.

<u>Condition (1)</u>: If P satisfies condition (0), and in addition, the generalized null vectors of maximal height occurring as columns of P^{-1} are orthogonal to all the generalized eigenvectors of A corresponding to non-zero eigenvalues, we say that P satisfies condition (1) with respect to A.

<u>Condition (2)</u>: If the null vectors of A appearing as columns of P^{-1} are orthogonal to all the generalized eigenvectors of A which are not null vectors of A, we say that P satisfies condition (2).

We note that for some matrices it is not possible to construct a matrix P such that P satisfies either condition (1) or condition (2), and at the same time transform the matrix into its Jordan canonical form.

Before establishing several properties for which the existence of a P satisfying the above conditions is sufficient to guarantee, we establish some subspace relationships.

Theorem 4.9: Let X be an n-dimensional Hilbert space, and A a matrix representation of a linear operator on X.

Then

$$\text{a)} \quad X = N(A^+) \oplus R(A),$$

$$\text{b)} \quad X = N(A) \oplus R(A^+).$$

Proof: To prove (a) it is sufficient to establish that any vector x in X can be written as $y + z$ where $y \in N(A^+)$ and $z \in R(A)$ and also that the intersection of $N(A^+)$ with $R(A)$ contains only the zero vector. It follows from equations (11) and (12) that AA^+ is a projection operator on $R(A)$ and A^+A is a projection operator on $R(A^+)$. It also follows that $(I - AA^+)$ is a projection operator on $N(A^+)$ and $(I - A^+A)$ is a projection operator on $N(A)$. Now, any vector x in X can be written as $x = AA^+x + (I - AA^+x)$ where $AA^+x \in R(A)$ and $(I - AA^+)x \in N(A^+)$. Assume x is in $N(A^+)$, then $AA^+x = 0$. If x is also in $R(A)$, then $AA^+x = x$. But this implies that $x = 0$. The proof of statement (b) follows similarly.

Theorem 4.10: If there exist a matrix P such that $PAP^{-1} = C$ where C is the Jordan canonical form of A, and P satisfies conditions (0) and (2), then

$$\text{a)} \quad R(A^+) = R(A^*),$$

$$\text{b)} \quad (A^+A)^* = A^+A,$$

$$\text{c)} \quad A^+AA^* = A^*,$$

$$\text{d)} \quad AA^+(A^+)^* = (A^+)^*,$$

$$\text{e)} \quad (AA^*)^+ = (A^*)^+A^+.$$

Proof: (a) Since $R(A^*) = N(A)^{\perp}$, it suffices to show that $R(A^+) = N(A)^{\perp}$. Let x be any vector in $R(A^+)$, then x is a linear combination of columns of P^{-1} which are not generalized null vectors of A^+ of maximal height, hence, not null vectors of A. But, the null vectors of A appearing as columns of P^{-1} form a basis for $N(A)$, and by condition (2) are orthogonal to all other generalized eigenvectors of A which are not null vectors of A. It follows that $R(A^+) = N(A)^{\perp}$.

(b) Since $R(A^+) = N(A)^{\perp}$, it is easily established that A^+A is an orthogonal projection operator on $R(A^+)$. But, this implies that $(A^+A)^* = A^+A$.

(c) Let $x \varepsilon X$, then $A^*x \varepsilon R(A^*) = R(A^+)$ so that $A^*x \varepsilon R(A^+)$. But, A^+A is an orthogonal projection operator on $R(A^+)$, hence $A^+AA^*x = A^*x$ from which it follows that $A^+AA^* = A^*$.

(d) Since $R(A^+) = R(A^*)$ we have that $R(A^{+*}) = R(A^{**}) = R(A)$. Let $x \varepsilon X$, then $A^{+*}x \varepsilon R(A)$. Since AA^+ is a projection operator on $R(A)$ we have that $AA^+A^{+*}x = A^{+*}x$. Hence $AA^+A^{+*} = A^{+*}$.

(e) Let $x_1, x_2, \ldots, x_n$ be a basis for X such that $x_1, x_2, \ldots, x_k$ spans $R[(AA^*)^+]$ and $x_{k+1}, \ldots, x_n$ spans $N(AA^*) = N(A^*)$. Then $AA^*x_1, \ldots, AA^*x_k$ spans $R(AA^*) = R(A)$. Estend this to a basis for X with $z_{k+1}, \ldots, z_n$, such that $z_{k+1}, \ldots, z_n$ spans $N[(AA^*)^+]$. Also $A^*x_1, \ldots, A^*x_k$ spans $R(A^*)$. Extend this to a basis for X with $y_{k+1}, \ldots, y_n$ such that $y_{k=1}, \ldots, y_n$ spans $R(A^*)^{\perp} = N(A)$. Using the fact that AA^+ and A^+A are projection operators we have the following:

$$A^{+}: \quad AA^{*}x_i \quad A^{*}x_i \quad \rightarrow \quad i = 1, 2, \ldots, k$$

$$z_i \quad 0 \quad \rightarrow \quad i = k + 1, \ldots, n$$

and

$$A^{*+}: \quad A^{*}x_i \quad x_i \quad \rightarrow \quad i = 1, 2, \ldots, k$$

$$y_i \quad 0 \quad \rightarrow \quad i = k + 1, \ldots, n$$

But

$$(AA^{*}): \quad AA^{*}x_i \quad x_i \quad \rightarrow \quad i = 1, 2, \ldots, k$$

$$z_i \quad 0 \quad \rightarrow \quad i = k + 1, \ldots, n$$

Hence, it follows that $(AA^{*})^{+} = A^{*+}A^{+}$.

<u>Theorem 4.11</u>: If there exist a matrix P such that $PAP^{-1} = C$ where C is the Jordan canonical form of A and P satisfies condition (1), then

a) $N(A^{+}) = N(A^{*})$,

b) $(AA^{+})^{*} = AA^{+}$,

c) $A^{*}AA^{+} = A^{*}$,

d) $(A^{*}A)^{+} = A^{+}A^{*+}$.

<u>Proof</u>: (a) It suffices to show that $N(A^{+}) = R(A)^{\perp}$, since $N(A^{*}) = R(A)^{\perp}$. Let x be any vector in $R(A)$, then x is a linear combination of columns of P^{-1} which are not generalized null vectors of A of maximal height, hence, not null vectors of A^{+}. By condition (1), the generalized null vectors of maximal height

of A are orthogonal to all other generalized eigenvectors of A which are not generalized null vectors of A of maximal height. Hence, it follows that $N(A^+) = R(A)^\perp$.

(b) Since $N(A^+) = R(A)^\perp$, AA^+ is an orthogonal projection operator on $R(A)$ which implies that $(AA^+)^* = AA^+$.

(c) Using (b) in the equation $AA^+A = A$ we have $(AA^+)^*A = A$. Taking the conjugate transpose of this equation we have that $A^*AA^+ = A^*$.

(d) Let $x_1, \ldots, x_n$ be a basis for X such that $x_1, \ldots, x_r$ spans $R[(A^*A)^+]$ and $x_{r+1}, \ldots, x_n$ spans $N(A^*A) = N(A)$. Then $A^*Ax_1, \ldots, A^*AX_r$ spans $R(A^*A) = R(A^*)$. Extend this to a basis for X with vectors $z_{r+1}, \ldots, z_n$ such that they span $N[(A^*A)^+]$. Now $Ax_1, \ldots, Ax_r$ spans $R(A)$. Extend this to a basis for X with vectors $y_{r+1}, \ldots, y_n$ such that they span $N(A^*) = N(A^+)$. Then using the fact that A^+A and AA^+ are projection operators we get:

$$\begin{array}{llcll} A^{*+} & : \; A^*Ax_i & \to & Ax_i & i = 1, 2, \ldots, r \\ & \quad z_i & \to & 0 & i = r+1, \ldots, n. \end{array}$$

and

$$\begin{array}{llcll} A^{+} & : \; Ax_i & \to & x_i & i = 1, 2, \ldots, r \\ & \quad y_i & \to & 0 & i = r+1, \ldots, n. \end{array}$$

But

$$\begin{array}{llcll} (A^*A)^{+} & : \; A^*Ax_i & \to & x_i & i = 1, 2, \ldots, r \\ & \quad z_i & \to & 0 & i = r+1, \ldots, n. \end{array}$$

It follows that $(A^*A)^+ = A^+A^{*+}$.

<u>Theorem 4.12</u>: Let $PAP^{-1} = C$ where C is the Jordan canonical form of A. Conditions (1) and (2) are necessary and sufficient for the following:

$$1)\quad (AA^{+})^{*} = AA^{+},$$

$$2)\quad (A^{+}A)^{*} = A^{+}A.$$

<u>Proof</u>: The sufficiency is established in Theorems 4.10 and 4.11. To show the necessity, assume P does not satisfy either condition (1) or condition (2). If P does not satisfy condition (1) we will show that $R(A) \neq N(A^{+})^{\perp}$. Let $x \in R(A)$, then x is a linear combination of columns of P^{-1} which are not generalized null vectors of maximal height for A. But, since P does not satisfy condition (1), there is a generalized null vector of maximal height for A, and hence a null vector of A^{+}, which is not orthogonal to $R(A)$. Hence $R(A) \neq N(A^{+})^{\perp}$.

But, $R(A) \neq N(A^{+})^{\perp}$ implies that AA^{+} is not an orthogonal projection operator on $R(A)$, which implies that $(AA^{+})^{*} \neq AA^{+}$.

In case P does not satisfy condition (2) we establish that $R(A^{+}) \neq N(A)^{\perp}$. Let $x \in R(A^{+})$, then x is a linear combination of columns of P^{-1} which are not generalized null vectors of A^{+} of maximal height, hence, not null vectors of A. But, since P does not satisfy condition (2), there is a null vector of A which is not orthogonal to $R(A^{+})$. Hence $R(A^{+}) \neq N(A)^{\perp}$. But this implies that $A^{+}A$ is not an orthogonal projection operator on $R(A^{+})$ which implies that $(A^{+}A)^{*} \neq A^{+}A$.

Thus, it follows that P satisfying conditions (1) and (2) is necessary and sufficient for the pseudo inverse defined above to be the same as the pseudo inverse defined by Penrose. A special case of Theorem 4.12 which is of interest in its own right is the following.

Corollary 4.2: If A is unitarily equivalent to C, its Jordan canonical form, then

1) $(A^+A)^* = A^+A,$

2) $(AA^+)^* = AA^+.$

Proof: If A is unitarily equivalent to C, then there exists a unitary matrix U such that $UAU^* = C$. Since the columns of U^* are mutually orthogonal we have conditions (1) and (2), and thus the conclusion by Theorem 4.13.

In particular, if A is normal the definitions are equivalent.

Theorem 4.13: If $\alpha \neq 0$, then $(\alpha A)^+ = \alpha^{-1}A^+$.

Proof: If $\alpha = 1$, the theorem is trivial, so assume that $\alpha \neq 1$. Let $P(\alpha A)P^{-1} = C$ where C is the Jordan canonical form of αA. Then $(\alpha A)^+ = P^{-1}C^IP$. Let C_1 be the direct sum of the Jordan blocks corresponding to non-zero eigenvalues, and C_2, C_3, . . ., C_k be the Jordan blocks corresponding to the zero eigenvalue. Then, without loss of generality we assume that $C = \text{diag}$ $(C_1, C_2, \ldots, C_k)$. Let Q_1 be the matrix such that $Q_1(\alpha^{-1}C_1)Q_1^{-1} = D_1$ where D_1 is the Jordan canonical form of $\alpha^{-1}C_1$. If C_i,

$i = 2, 3, \ldots, k$, is an n_i by n_i matrix, let $Q_i = \text{diag}(\alpha^{n_i-1}, \alpha^{n_i-2}, \ldots, \alpha, 1)$. It is easily verified by direct multiplication that $Q_i(\alpha^{-1}C_i)\, Q_i^{-1} = D_i$ where $D_i = C_i$. Letting $Q = \text{diag}(Q_1, Q_2, \ldots, Q_k)$, and $D = \text{diag}(D_1, D_2, \ldots, D_k)$ it follows that $Q\alpha^{-1}CQ^{-1} = D$ where D is the Jordan canonical form of $\alpha^{-1}C$. Then $(\alpha^{-1}C)^+ = Q^{-1}D^IQ$. From the form of Q, it is easily verified that it satisfies both conditions (1) and (2). Hence, by Theorem 4.12 $(\alpha^{-1}C)^+ = (\alpha^{-1}C)^I$. But Price [73] has shown that $(\alpha^{-1}C)^I = \alpha C^I$. Now, $PAP^{-1} = \alpha^{-1}C = Q^{-1}DQ$ implies that $QPAP^{-1}Q^{-1} = D$. It is easily shown that QP satisfies condition (0). Thus, $QPA^+P^{-1}Q^{-1} = D^I$ or $PA^+P^{-1} = Q^{-1}D^IQ$.

$$\begin{aligned}
\text{Hence,}\quad (\alpha A)^+ &= P^{-1}C^IP \\
&= P^{-1}[\alpha^{-1}(\alpha^{-1}C)^I]\, P \\
&= \alpha^{-1}P^{-1}(\alpha^{-1}C)^+\, P \\
&= \alpha^{-1}P^{-1}(Q^{-1}D^IQ)\, P \\
&= \alpha^{-1}P^{-1}(PA^+P^{-1})\, P \\
&= \alpha^{-1}A^+ .
\end{aligned}$$

Lemma 4.4: If there exist a matrix P such that $PAP^{-1} = C$ and P satisfies conditions (0) and (2), then $(P^*)^{-1}$ satisfies condition (1) provided that the null vectors of A appearing as columns of P^{-1} are mutually orthogonal.

Proof: Let x_i, $i = 1, 2, \ldots, n$, be the columns of P^{-1}, and y_i, $i = 1, 2, \ldots, n$ be the columns of P^*. Now $\{x_i\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$ form biorthogonal bases for X. Let x_i

be a null vector for A. Expanding x_i in terms of the y_i, we get

$$x_i = \sum_{j=1}^{n} (x_i, x_j)\, y_j$$

But, $(x_i, x_j) = 0$ unless $i = j$, since P satisfies condition (2), and the null vectors of A appearing as columns of P^{-1} are mutually orthogonal. Then $(x_i, y_j) = (x_i, x_i)(y_i, y_j)$ for any $j = 1, 2, \ldots, n$. But $(x_i, y_j) = 0$ for $j \neq i$, hence $(y_i, y_j) = 0$ for $j \neq i$. Now, since x_i is a null vector of A, it follows that y_i is a generalized null vector of maximal height for A^*. Hence $(P^*)^{-1}$ satisfies condition (1).

Theorem 4.14: If there exist a matrix P such that $PAP^{-1} = C$ where C is the Jordan canonical form of A and if P satisfies the conditions in Lemma 4.4, then

a) $(A^+)^* = (A^*)^+$,

b) $A^{*+}A^*A = A$,

c) $A^+A^{+*}A^* = A^+$.

Proof: (a) Let $PAP^{-1} = C$ where C is the Jordan canonical form of A. Let C_1 be the direct sum of the Jordan blocks corresponding to non-zero eigenvalues, and $C_2, C_3, \ldots, C_k$ be the Jordan blocks corresponding to the zero eigenvalue. Without loss of generality we assume that $C = \text{diag}\,(C_1, C_2, \ldots, C_k)$. Hence, $C^* = \text{diag}\,(C_1^*, C_2^T, \ldots, C_k^T)$. Let Q_1 be the matrix such that $Q_1C_1^*Q_1^{-1} = D_1$ where D_1 is the Jordan canonical form of C_1^*. If C_i^T, $i = 2, 3, \ldots, k$

is an n_i by n_i matrix, let Q_i be the n_i by n_i matrix all of whose elements are zero except the elements on the diagonal from the $(1, n_i)$ position to the $(n_i, 1)$ position. Each of the elements on this diagonal is one. That is, in this case Q_i is a permutation matrixmatrix which reverses the order of the columns of a matrix when the matrix is multiplied on the right by Q_i, and reverses the order of the rows of the matrix when the matrix is multiplied on the left by Q_i. It is evident that $Q_i^{-1} = Q_i$ and $Q_i C_i^T Q_i^{-1} = D_i$ where $D_i = C_i$, i= 2, 3, . . ., k. Letting $D = \text{diag}\,(D_1, D_2, \ldots, D_k)$ and $Q = \text{diag}\,(Q_1, Q_2, \ldots, Q_k)$ we have $QC^*Q^{-1} = D$ where D is the Jordan canonical form of C^*. Now, $PAP^{-1} = C$ implies that $P^{*-1}A^*P^* = C^* = Q^{-1}DQ$. This implies that $QP^{*-1}A^*P^*Q^{-1} = D$. If y is a generalized null vector of maximal height for A^* which appears as a column of P^*Q^{-1}, it follows from the form of Q and y is a column of P^* which is of maximal height for A^*. But, by Lemma 4.4, $(P^*)^{-1}$ satisfies condition (1) which implies condition (0). Hence $Q(P^*)^{-1}$ satisfies condition (0). Thus, we have $QP^{*-1}(A^*)^+P^*Q^{-1} = D^I$. Now Q obviously satisfies condition (0) with respect to C^* so that $(C^*)^+ = Q^{-1}D^IQ$. But, Q also satisfies conditions (1) and (2) so that $(C^*)^+ = (C^*)^I$, and Desoer and Whalen [36] have shown that $(C^*)^I = (C^I)^*$. Hence,

$$\begin{aligned}
(A^*)^+ &= P^*(Q^{-1}D^IQ)P^{*-1} \\
&= P^*(C^*)^+P^{*-1} \\
&= P^*(C^I)^*P^{*-1} \\
&= (A^+)^*.
\end{aligned}$$

(b) Since P satisfies conditions (0) and (2), $R(A^+) = R(A^*)$ by Theorem 4.10(a). This implies that $R(A^{+*}) = R(A^{**}) = R(A)$. For any vector x in X, it follows that $Ax \in R(A^{+*})$. Since $(A^*)^+A^*$ is a projection operator on $R(A^{*+}) = R(A^{+*})$, it follows that $(A^*)^+A^*Ax = Ax$ and thus $(A^+)^*A^*A = A$.

(c) Since $R(A^{+*}) = R(A)$, it follows that $AA^+A^{+*} = A^{+*}$. Taking conjugate transposes we get $A^+A^{+*}A^* = A^+$.

<u>Definition 4.2:</u> The annihilator S^o of any subset S of X, is the set of all vectors y in the dual space of x, say X^*, such that (x, y) is identically zero for all x in S.

<u>Theorem 4.15:</u> $(A^+A)^*$ and $(AA^+)^*$ are projection operators on the spaces of annihilators of $N(A)$ and $N(A^+)$, respectively.

<u>Proof:</u> Using the fact that $A^+ = A^+AA^+$ we have $(A^+A)^* = (A^+AA^+A)^* = (A^+A)^*(A^+A)^*$. Hence $(A^+)^*$ is idempotent, and thus a projection operator. From Theorem 4.9, $X = N(A) \oplus R(A^+)$, and hence that $X^* = N(A)^o \oplus R(A^+)^o$, where Z^o is the space of annihilators of Z. Let y be a vector in $N(A)^o$, then for any x in X we have $(x, y) = (A^+Ax + [I - A^+A]\, x, y) = (A^+Ax, y) + ([I - A^+A]\, x, y)$. But, $I - A^+A$ is a projection operator on $N(A)$ so that $(I - A^+A)\, x \in N(A)$. Hence, $([I - A^+A]\, x, y) = 0$ so that $(x, y) = (A^+Ax, y) = (x, (A^+A^*\, y)$. Since this must hold for each x in X, this implies that $(A^+A)^*\, y = y$. Now assume $y \in R(A^+)^o$. Then for any x in X we have that $(x, y) = (A^+Ax, y) + ([I - A^+A]x, y) = ([I - A^+A]x, y)$ since $A^+Ax \in R(A^+)$. Now $([I - A^+A]x, y) = (x, [I - A^+A]^*\, y) = (x, y) - x, (A^+A)^*\, y)$.

Hence $(x, (A^+A)^* y) = 0$. Again, since this must hold for each x in X, it follows that $(A^+A)^* y = 0$. The other part follows in a similar manner.

Theorem 4.16: $R(A)^\circ$ is invariant under A^* and $R(A^+)^\circ$ is invariant under $(A^*)^+$.

Proof: Let $y \varepsilon R(A)^\circ$, then for any x in X we have $(x, A^* y) = (Ax, y) = 0$ since $Ax \varepsilon R(A)$. This implies that $A^* y \varepsilon R(A)^\circ$ and thus that $R(A)^\circ$ is invariant under A^*. The second part of the theorem follows similarly.

Theorem 4.17: Let $tr(A)$ represent the trace of the matrix A. Then $tr(A^+A) = tr(AA^+) = r(A) = r(A^+)$.

Proof: The first equality is a property of the trace. The last one was established in Theorem 4.6. We now show that $tr(AA^+) = r(A)$. Using the properties of the trace we have that $tr(AA^+) = tr(P^{-1}CC^IP) = tr(CC^IPP^{-1}) = tr(CC^I)$. But, Penrose [70] has shown that $tr(CC^I) = r(C)$, hence $tr(AA^+) = r(C) = r(A)$.

We proceed now to explore the use of A^+ in solving systems of linear equations. A method for computing A^+ is given and an example is presented.

The following theorem is essentially a consequence of a minimal property for the Penrose pseudo inverse.

Theorem 4.18: If the system of linear equations $Ax = b$ is consistent, then $x_1 = A^{+}b$ is a solution of the system. If $Ax = b$ is inconsistent, then for any x in X,

$$||P(Ax_1 - b)|| \leq ||P(Ax - b)|| \; , \tag{21}$$

where $A = P^{-1}CP$ and $x_1 = A^{+}b$.

Proof: It suffices to prove the inequality (21). For if there exist a vector x_2 in X such that $Ax_2 = b$, $||P(Ax_2 - b)|| = 0$. Thus, if (21) is valid, $||P(Ax_1 - b)|| = 0$, which implies that $Ax_1 = b$.

Now $P(Ax_1 - b) = P(AA^{+}b - b) = PAA^{+}b - Pb$. Using the definition of A^{+}, $P(Ax_1 - b) = CC^{I}Pb - Pb$. Let $Pb = b_1 + b_2$, where $b_1 \varepsilon R(C)$ and $b_2 \varepsilon R(C)^{\perp}$. Then $P(Ax_1 - b) = CC^{I}(b_1 + b_2) - (b_1 + b_2) = b_1 - b_1 - b_2 = -b_2$ since CC^{I} is a projection on $R(C)$. Thus

$$||P(Ax_1 - b)|| = ||b_2||. \tag{22}$$

If x is any vector belonging to X, then $P(Ax - b) = CPx - Pb$. Since $CPx \varepsilon R(C)$,

$$||CPx - (b_1 + b_2)|| = ||b_2|| + ||CPx - b_1||. \tag{23}$$

The inequality (21) follows immediately from (22) and (23).

Since A^{+} is unique, the solution $A^{+}b$ to the consistent system of equations $Ax = b$ is unique. One might ask from whence comes this uniqueness. Lanczos [60] has given an indirect answer to this question. In fact, his answer applies directly to the solution of such a set of equations if the Penrose definition of the pseudo inverse is used. Lanczos shows that the uniqueness for the solution is obtained by adding conditions to be satisfied by the solution vector x. The pseudo inverse defined by Penrose yields the unique solution obtained by adding the condition which together with $Ax = b$ has the unique solution $A^{+}b$.

As has been pointed out before, the columns of P^{-1} are a canonical basis for A. Also, the columns of P^{*} are eigenvectors or generalized eigenvectors of A^{*}. Now if y_j is a generalized null vector of maximal height of A^{*} which appears as a column of P^{*}, then the j-th column, x_j, of P^{-1} is a null vector of A and, consequently, a generalized null vector of maximal height of A^{+}. But if $x \in R(A^{+})$, then x is a linear combination of columns of P^{-1} which are not null vectors of A. This means that, if X is the set of all generalized null vectors of maximal height of A^{*} which appear as columns of P^{*} and X_1 is the linear span of X, then $R(A^{+}) = X_1$. Thus, if $x_1 = A^{+}b$, then x_1 satisfies the system

$$\begin{bmatrix} A \\ \hat{G} \end{bmatrix} x_1 = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{24}$$

where $\hat{G}^{*}$ is a matrix whose columns span X_1. The fact that the columns of $\hat{G}^{*}$ span X_1 implies that the columns of $\hat{G}^{*}$ do not belong to $R(A^{*})$. Hence $\binom{A}{\hat{G}}$ has rank n. Hence, the auxiliary condition is that $\hat{G}x_1 = 0$ or in words, x_1 must be orthogonal to all generalized null vectors of maximal height of A^*.

To obtain a computational procedure for obtaining A^+, we begin with a factorization of A. Let B be an n by r matrix of rank r such that $A = BG$. Now $B_1 = (B^*B)^{-1}B^*$ is a left inverse of B and $G_1 = G^*(GG^*)^{-1}$ is a right inverse of G. Making use of equation (11), we have

$$GA^+B = I_r \tag{25}$$

where I_r is the r by r identity matrix. Let $\hat{B}$ be an n-r by n matrix such that if x is a column of $\hat{B}$, then x is a generalized null vector of A of maximal height, say k, and x is orthogonal to all generalized null vectors of height at most k. In other words, the columns of $\hat{B}$ could be chosen as columns of P^{-1} which are generalized null vectors of maximal height for A. Then the partitioned n by n matrix $[B, \hat{B}]$ has rank n. Furthermore, since a generalized null vector of maximal height for A is a null vector for A^+ we have

$$A^+\hat{B} = 0. \tag{26}$$

From the discussion given above, it follows that

$$\hat{G}A^+ = 0\ . \tag{27}$$

The matrix $\begin{pmatrix} G \\ \hat{G} \end{pmatrix}$ is an n by n matrix of rank n. Combining equations (25), (26) and (27) we have

$$\begin{bmatrix} G \\ \hat{G} \end{bmatrix} A^{+} [B, \hat{B}] = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

and solving for A^{+} we get

$$A^{+} = \begin{bmatrix} G \\ \hat{G} \end{bmatrix}^{-1} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} [B, \hat{B}]^{-1} \quad (28)$$

In view of equation (28) the first step in the computation of A^{+} is the determination of G and G. Theorem 4.1 gives us a technique for obtaining the generalized null vectors of A of maximal height. These orthonormal vectors are columns of $\hat{B}$. $\hat{G}$ is obtained by computing the appropriate vectors in the dual chains.

We compute a simple example to illustrate the technique. Consider the matrix

$$A = \begin{bmatrix} 4 & 2 & -3 \\ -2 & 0 & 1 \\ 2 & 2 & -2 \end{bmatrix}$$

The rank of A is 2. A null vector of A is $x_1 = (1, 1, 2)^T$. We solve the system of equations

$$\begin{bmatrix} A^2 \\ x_1^T \end{bmatrix} x_2 = 0$$

to obtain $x_2 = (2, -4, 1)^T$. The rank of A^2 is one which is the same as the rank of A^3. So we conclude that x_2 is a generalized null vector of maximal height 2 for A and the only column of $\hat{B}$. Now $Ax_2 = (-3, -3, -6)^T = x_3$. The single row of $\hat{G}$ is the transpose of the solution of

$$\begin{bmatrix} (A^*)^2 \\ x_2^T \\ x_3^T \end{bmatrix} y = \begin{bmatrix} \phi \\ 0 \\ 0 \end{bmatrix}$$

where ϕ is the three by one null vector. This solution is $y = (-1/18, -1/18, -1/9)^T$. Let

$$B = \begin{bmatrix} 2 & 1 \\ -1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \end{bmatrix}$$

Then

$$A^+ = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1/18 & -1/18 & -1/9 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ -1 & 0 & -4 \\ 1 & 1 & 1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 5/12 & -1/12 & -3 \\ -1/12 & 5/12 & -3 \\ -1/6 & -1/6 & -6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 4/3 & 1/3 & -4/3 \\ -1 & 0 & 2 \\ -1/3 & -1/3 & 1/3 \end{bmatrix}$$

$$= 1/36 \begin{bmatrix} 23 & 5 & -26 \\ -19 & -1 & 34 \\ -2 & -2 & -4 \end{bmatrix}$$

4.3. Pseudo Inverses of Non-Square Matrices

We now proceed to investigate pseudo inverses of not necessarily square matrices on finite dimensional Hilbert spaces. Let X and Y be finite dimensional Hilbert spaces of dimension m and n, respectively. Let A be a linear transformation from X into Y. We will represent the set of all linear transformations from X into Y by $[X, Y]$.

Definition 4.3: Let $A \varepsilon [X, Y]$. A pseudo inverse of A will mean a linear transformation $A^+ \varepsilon [Y, X]$ such that

$$AA^+A = A, \tag{1}$$

$$A^+AA^+ = A^+ \tag{2}$$

Theorem 4.19: Let $A \varepsilon [X, Y]$

a) If $B \varepsilon [Y, X]$ such that $ABA = A$ and $BAB = B$, then $X = N(A) \oplus R(B)$ and $Y = R(A) \oplus N(B)$.

b) Conversely, if U and V are subspaces of X and Y, respectively, such that $X = N(A) \oplus U$ and $Y = R(A) \oplus V$, then there exists a unique B such that $B \varepsilon [Y, X]$ and $ABA = A$, $BAB = B$ with $R(B) = U$, $N(B) = V$.

Proof: (a) Let $x \varepsilon X$. Then x can be written as $BAx + (I - BA)x$ where $BAx \varepsilon R(B)$ and $(I - BA)x \varepsilon N(A)$. Assume x is in $N(A)$ and also in $R(B)$. Then $Ax = 0$, and there exist a vector y in Y such that $By = x$. Hence $ABy = 0$ which implies that $BABy = 0$.

Now $BAB = B$ so that $By = x = 0$. Hence $X = N(A) \oplus R(B)$. The fact that $Y = R(A) \oplus N(B)$ follows similarly.

(b) To show this part, let U and V be as required. Let $x_1, x_2, \ldots, x_t, x_{r+1}, \ldots, x_n$ be a basis for X such that $x_1, x_2, \ldots, x_r$ spans U, and $x_{r+1}, x_{r+2}, \ldots, x_n$ spans $N(A)$. Then $Ax_1, Ax_2, \ldots, Ax_r$ spans $R(A)$. Choose $y_{r+1}, y_{r+2}, \ldots, y_m$ so that $Ax_1, Ax_2, \ldots, Ax_r, y_{r+1}, \ldots, y_m$ is a basis for Y. Define B as follows:

$$B(Ax_i) = x_i \qquad i = 1, 2, \ldots, r$$

$$By_i = 0 \qquad i = r + 1, \ldots, m$$

By Paige and Swift [69] this determines B uniquely. It follows from the construction of B that $N(B) = V$ and $R(B) = U$. Also, it follows immediately that $ABA = A$ and $BAB = B$.

In view of Theorem 4.19, any conditions which are sufficient to determine a unique pseudo inverse are equivalent to a specification of the null space and range of the pseudo inverse in question. We note that for the Penrose pseudo inverse, A^I, that $N(A^I) = R(A)^\perp$ and $R(A^I) = N(A)^\perp$. Also, for the definition given in (3) we note that $R(A^+) = X_1$ where X_1 is the linear span of the set of all generalized null vectors of maximal height of A^* which appear as columns of P^*, where of course, P satisfies condition (0). Also $N(A^+)$ is the space spanned by the set of generalized null vectors of maximal height of A which appear as columns of P^{-1}.

Definition 4.4: Let $X = S \oplus T$. If $x = s + t$ is an element of X, the linear transformation mapping X onto S such that s is the image of x under this transformation is called the projection of X onto S along T and will be denoted by $P_{S/T}$. If $S = T$, then it is an orthogonal projection and will be denoted by P_S.

Theorem 4.20: Let $X = S \oplus T$. Then P_S satisfies the following:

1) P_S exists and is unique
2) P_S is linear
3) $P_S \; P_S = P_S$
4) P_S is an orthogonal projection if and only if $P_S = P_S^* = P_S^2$.

Proof: These are well known results and are included for completeness. They may be found in Paige and Swift [69].

Theorem 4.21: Let $A \in [X, Y]$ and $B \in [Y, X]$ with B a pseudo inverse of A. Then

1) $AB = P_{R(A)/N(B)}$

2) $BA = P_{R(B)/N(A)}$

Proof: We establish (1), and (2) is established in a similar manner. Let x be any vector in X, then $x = y + z$, where $y \in R(A)$ and $z \in N(B)$. Since $ABA = A$ we have that $(AB)(AB) = AB$ so that AB is a projection. Also $ABx = ABy + ABz = ABy$. But

$y \in R(A)$, hence there exist a vector u such that $Au = y$. Therefore, $ABx = ABAu = Au = y$. Hence $AB = P_{R(A)/N(B)}$.

Now, if $A \in [X, Y]$ and $B \in [Y, X]$ with $R(A) = N(B)$ and $N(A) = R(B)$, then B is the Penrose pseudo inverse of A. For the remainder of this chapter, we will designate this unique pseudo inverse by A^I.

Theorem 4.22: Let $A \in [X, Y]$, $A^{\#} = A^I AB$ and $A^S = BAA^I$, where B is any pseudo inverse of A. Then $A^{\#}$ and A^S are pseudo inverses of A.

Proof: By direct substitution into the equations defining a pseudo inverse for A, and using the fact that B satisfies these equations we get

$$AA^{\#}A = AA^{+}ABA = AA^{+}A = A.$$

Also $A^{\#}AA^{\#} = A^{+}ABAA^{+}AB = A^{+}AA^{+}AB = A^{+}AB = A^{\#}$. Similarly, A^S is shown to be a pseudo inverse of A.

A more general result is given in this next theorem.

Theorem 4.23: If B and C are any pseudo inverses of A, then $A^S = BAC$ is a pseudo inverse of A.

Proof: By direct substitution again we get

$$AA^SA = ABACA = ACA = A$$

and

$$A^SAA^S = BACABAC = BABAC = BAC = A^S.$$

Hence BAC is a pseudo inverse of A.

This seems to indicate a relationship between any two pseudo inverses of A. The existing relationship is established in the next theorem.

Theorem 4.24: Let B and C be any two pseudo inverses of A, then there exist nonsingular P and Q such that

$$C = PBQ^{-1}$$

Proof: Let $x_1, x_2, \ldots, x_r, x_{r+1}, \ldots, x_n$ be a basis for X such that $x_1, x_2, \ldots, x_r$ spans R(B) and $x_{r+1}, x_{r+2}, \ldots, x_n$ spans N(A). Now $Ax_1, \ldots, Ax_r$ spans R(A). Complete this to a basis for Y by selecting $y_{r+1}, \ldots, y_n$ so that they span N(B). This completely defines B. Likewise, let $x'_1, x'_2, \ldots, x'_r, x'_{r+1}, \ldots, x'_n$ be a basis for X such that $x'_1, x'_2, \ldots, x'_r$ spans R(C) and $x'_{r+1}, \ldots, x'_n$ spans N(A). Now $Ax'_1, \ldots, Ax'_r$ spans R(A). Complete this to a basis for Y with $y'_{r+1}, \ldots, y'_m$, so that $y'_{r+1}, \ldots, y'_m$ spans N(C). Define P and Q as follows:

$$Px_i = x'_i \qquad i = 1, 2, \ldots, n$$

$$Q(Ax_i) = Ax'_i \qquad i = 1, 2, \ldots, r$$

$$Qy_i = y'_i \qquad i = r + 1, \ldots m$$

It follows from the above that $A = Q^{-1}AP = QAP^{-1}$ and that $C = PBQ^{-1}$.

Corollary 4.3: Let B be a pseudo inverse of A, then PBQ^{-1} is a pseudo inverse of A if and only if $Q^{-1}AP$ is a pseudo inverse of B.

Proof: Let PBQ^{-1} be a pseudo inverse of A. This implies that $PBQ^{-1}APBQ^{-1} = PBQ^{-1}$ and $APBQ^{-1}A = A$. Hence $(Q^{-1}AP)\,B(Q^{-1}AP) = Q^{-1}(APBQ^{-1})P = Q^{-1}AP$ and $B(Q^{-1}AP)B = P^{-1}(PBQ^{-1}APBQ^{-1})Q = P^{-1}(PBQ^{-1})Q = B$.

Hence $Q^{-1}AP$ is a pseudo inverse of B. The converse is established in a similar manner.

Consider the system of linear equations given by $Ax = b$, where x and b are vectors. A necessary and sufficient condition that a solution exist is that b is in the range of A. In case $b \notin R(A)$, the least squares solution is given by $x = A^I b$ where A^I is the Penrose pseudo inverse. However, if B is any pseudo inverse of A, and $A^\# = BAA^I$, then $x = A^\# b$ is also a least squares solution. This is the conclusion of the next theorem.

Theorem 4.25: Let $A^\# = BAA^I$ where B is any pseudo inverse of A, then $x_1 = A^\# b$ is a least squares solution to the system of linear equations given by $Ax = b$.

Proof: Consider $||Ax_1 - b||$. Substituting in for x_1 we get $||Ax_1 - b|| = ||AA^\# y - y|| = ||ABAA^I y - y|| = ||AA^I y - y||$.

The result follows from the work of Penrose [70].

We also note that the following theorems hold for any pseudo inverses, not necessarily just the Penrose pseudo inverse.

Theorem 4.26: For the matrix equation $AXB = C$ to have a solution X, a necessary and sufficient condition is that $AA^\# CB^\# B = C$ in which case, the general solution is

$$X = A^\# CB^\# + Y - A^\# AYBB^\#$$

where $A^{\#}$, $B^{\#}$ are any pseudo inverses of A and B respectively, and Y is arbitrary to within having the dimensions of X.

Proof: If X_1 satisfies $AXB = C$, then

$$C = AX_1B = AA^{\#}AXBB^{\#}B = AA^{\#}CB^{\#}B.$$

Conversely, if $C = AA^{\#}CB^{\#}B$, then $A^{\#}CB^{\#}$ is a particular solution. For the general solution, $AXB = 0$ must be solved. Any expression of the form $X = Y - A^{\#}AYBB^{\#}$ is a solution of $AXB = 0$. Hence the conclusion follows.

Theorem 4.27: A necessary and sufficient condition for the equations $AX = C$ and $XB = D$ to have a common solution is that each have a solution and $AD = CB$.

Proof: If $AX = C$ and $XB = D$ have a common sulution then clearly each has a solution and

$$AXB = BC$$

$$AXB = AD$$

so that $CB = AD$. In order to obtain the sufficiency of the condition, let

$$X = A^{\#}C + DB^{\#} - A^{\#}ADB^{\#}$$

where $A^{\#}$ and $B^{\#}$ are any pseudo inverses of A and B, respectively. It is easily verified by direct substitution that this is a solution provided $AD = CB$, $AA^{\#}C = C$ and $DB^{\#}B = D$.

CHAPTER 5

APPLICATIONS

5.1 Linear Systems of Equations:

We first consider the system $Ax = y$, where A is a p by n matrix of constants, x is an n by 1 vector of unknowns and y is a p by 1 vector of constants. There is no easy way to decide whether this system is consistent. The following is a simple technique using the Rao definition 2.4 of a pseudo inverse A^- of A to check for consistency and once consistency is established the solution is immediate.

Lemma 5.1: Let $A^-A = H$ for a given pseudo inverse A^-. Then

a) $H^2 = H$

b) $AH = A$

c) The solutions of $Ax = 0$ can be expressed as $(H - I)z$ where z is arbitrary.

d) A general solution of $Ax = y$, when consistent, is $A^-y + (H - I)z$.

e) q^*x has a unique value for all x satisfying the equations $Ax = y$, if $q^*H = q^*$.

Proof: a) Since, by theorem 2.5, $AA^-A = A$, premultiplying by A^- gives $A^-AA^-A = A^-A$ or $H^2 = H$.

b) Also, by Theorem 2.5, $A(A^-A) = A$ which implies that $AH = A$.

c) Since $AH = A$, $r(H) \geq r(A)$, where $r(\cdot)$ is the rank of $(\cdot)$. But, $A^{-}AH = H$ so that $r(H) \geq r(A)$. Hence, $r(H) = r(A)$. Also $r(H - I) = n - r(H) = n - r(A)$. Since $A(H - I) = 0$, the columns of $H - I$ supply all the solutions of $Ax = 0$. Hence, a general solution is $(H - I)z$ where z is arbitrary.

d) Since $A^{-}y$ is a particular solution of $Ax = y$, the general solution is $A^{-}y + (H - I)z$.

e) Substituting in $q^{*}x$ a general solution of $AX = y$ we get $q^{*}[A^{-}y + (H - I)z] = q^{*}A^{-}y + q^{*}Hz - q^{*}Iz = q^{*}A^{-}y$ if $q^{*}H = q^{*}$.

To avoid complications make A square by adding zeroes. Recall that given a matrix A, there exist a non-singular B such that $BA = H$ where H has the following properties.

a) The diagonal elements are 0 or 1.

b) If the i^{th} diagonal element is 1, all elements in the i^{th} column and all elements preceding 1 in the i^{th} row are 0.

c) If the j^{th} diagonal element is 0, all elements in the j^{th} row are 0, and also those below the 0 diagonal element in the j^{th} column.

Define the matrix G as a diagonal matrix with its i^{th} diagonal element 1 if the i^{th} diagonal element of H is 0, and 0 otherwise.

Theorem 5.1: With A, B, H and G as defined above the following are true:

a) $H^2 = H$ (H is idempotent)

b) $AH = A$

c) $ABA = A$ and hence B is a pseudo inverse of A by definition (2.4).

d) $Ax = y$ is consistent if and only if $GBy = o$, i.e., if the r_1^{th}, r_2^{th}, . . ., rows of H are null, then the r_1^{th}, r_2^{th}, . . ., elements in By must be 0.

e) A general solution of $Ax = y$ is $By + (H - I)z$ where z is arbitrary.

f) $q^* x$ is unique if and only if when x satisfies $Ax = y$ we have $q^* H = q^*$.

Proof: a) This is established by direct multiplication.

b) Since $BA = H$ and B is non-singular, we have that $A = B^{-1}H$. Hence, $AH = B^{-1}H^2$ and by (1), $B^{-1}H^2 = B^{-1}H = A$. Thus, $AH = A$.

c) Since $BA = H$, $ABA = AH = A$.

d) Since B is non-singular, if $Ax = y$ is consistent, so is $BAx = By$ or $Hx = By$, and conversely. If the r^{th} row of H is zero, then the r^{th} element of Hx is zero and so must be the r^{th} element of By. Conversely, if this is true $x = By$ is obviously a solution of $Hx = By$.

e) and f) are established as in lemma 5.1.

If, in addition to B, we know which of the rows of H are null, we have an automatic test for consistency of $Ax = y$, while finding a solution. Let the $r_1^{th}, r_2^{th}, \ldots, r_k^{th}$ rows in H be null. Then we need only compute By and examine whether the $r_1^{th}, \ldots, r_k^{th}$ elements are 0. If they are, the equations are consistent, in which case By itself is a solution. It is important to note that B is a non-singular inverse, although A may be singular. This is necessary for the consistency test given in (d). The Penrose pseudo inverse is necessarily singular if A is singular, but if the system is known to be consistent, it can be used to obtain a general solution of $Ax = y$, as is ascertained in the next theorem.

Theorem 5.2: The general solution of the vector equation $Ax = y$ is $x = A^+y + (I - A^+A)z$, where z is arbitrary, provided that the equation has a solution.

Proof: Suppose x satisfies $Ax = y$. Then

$$\begin{aligned} y = Ax &= AA^+y + A(I - A^+A)z \\ &= AA^+y + (A - AA^+A)z \\ &= AA^+y \quad \text{since} \quad A = AA^+A. \end{aligned}$$

Hence, A^+y is a particular solution of $Ax = y$. For the general solution, we must solve $Ax = 0$. Now any expression of the form $x = (I - A^+A)z$ satisfies $Ax = 0$ and conversely if $AX = 0$ then x can be expressed in the form $(I - A^+A)z$.

We now consider more general systems of linear equations in the next two theorems.

Theorem 5.3: A necessary and sufficient condition for the equation $AXB = C$ to have a solution is

$$AA^+CB^+B = C,$$

in which case the general solution is

$$X = A^+CB^+ + Y - A^+AYBB^+,$$

where Y is arbitrary to within having the dimensions of X.

Proof: Suppose X satisfies $AXB = C$. Then $C = AXB = AA^+AXBB^+B = AA^+CB^+B$. Conversely, if $C = AA^+CB^+B$, then A^+CB^+ is a particular solution of $AXB = C$. For the general solution we must solve $AXB = 0$. Now any expression of the form $X = Y - A^+AYBB^+$ satisfies $AXB = 0$ and conversely, if $AXB = 0$, then $X = X - A^+AXBB^+$. It follows that the general solution is as given.

It might be noted that the only property required of A^+ for Theorem 5.3 is $AA^+A = A$.

Theorem 5.4: A necessary and sufficient condition for the equations $AX = C$, $XB = D$ to have a common solution is that each equation should individually have a solution and that $AD = CB$.

Proof: The condition is obviously necessary. To show that it is sufficient, put $X = A^+C + DB^+ - A^+ADB^+$, which is a solution if the required conditions $AA^+C = C$, $DB^+B = D$, $AD = CB$ are satisfied. The first two conditions come from Theorem 5.3 to guarantee that each equation individually has a solution. Again it should be noted that the only property required of A^+ is that $AA^+A = A$.

When the system $Ax = y$ does not admit of an exact solution, $x = A^+y + (I - A^+A)z$ as given in Theorem 5.2 nevertheless gives a "best" solution in the sense of least squares. That is, if y is a vector which is not in the range (column) space of A, $x = A^{+}y + (I - A^{+}A)z$ is a vector such that Ax is the projection of y on that space. Thus Ax is as "close" to y as it can be made, or, in other words, the sum of the squares of the residuals is a minimum. This is the conclusion of the next theorem.

Theorem 5.5: Let y be any n by 1 vector and $x_1 = A^{+}y$, where A is an n by p matrix. Then

$$||Ax_1 - y|| \leq ||Ax - y|| \quad \text{for any } p\text{-vector } x,$$

and

$$||x_1|| \leq ||x_0|| \quad \text{for all } x_0 \text{ satisfying the above}$$

inequality.

Proof: Let $y = y_1 + y_2$ with $y_1 \varepsilon R(A)$ and $y_2 \varepsilon R(A)$. Then $||Ax_1 - y|| = ||AA^{+}y - y|| = ||y_1 - y|| = ||y_2||$. On the other hand, for any p-vector x, let $Ax = y_3$. Certainly $y_3 \varepsilon R(A)$, thus $||Ax - y||^2 = ||y_3 - y_1 - y_2||^2 = ||y_3 - y_1||^2 + ||y_2||^2$. The last equality follows since the vector $y_3 - y_1$ is orthogonal to y_2. Hence, the desired inequality follows. Any vector x_0 satisfying $||Ax_1 - y|| \leq ||Ax_0 - y||$ is of the form $x_1 + x_2$ where x_2 is orthogonal to x_1. Hence,

$||x_0|| = ||x_1|| + ||x_2||$ from which we obtain $||x_1|| \leq ||x_0||$.

Note that x_1 is not only the least squares solution, which may not be unique, but also the vector of minimum norm which is a least squares solution and thus x_1 is unique.

5.2 Distribution Theory:

If x is a column vector of n random variables which have a joint n-dimensional Gaussian (or normal) distribution with mean vector m and covariance matrix V, we denote it as $x \sim N(m, V)$. In this, if $V = I$, then $y = \sum_{i=1}^{m} x_i^2 = x^T x$ has a known distribution, called the noncentral chi-square, and this is written as $y \sim \chi^2(n, \lambda)$, where the so-called noncentrality parameter $\lambda = 1/2 m^T m$. If $\lambda = 0$, the noncentral chi-square is the central chi-square.

Theorem 5.6: Let the $p \times 1$ random vector $x \sim N(0, V)$, where $r(V) = k \leq p$. A necessary and sufficient condition that a quadratic form $x^T A x$ has a χ^2 distribution is that V is a pseudo inverse of A by definition (2.4).

Proof: The result is well known when V is nonsingular. In any case V can be written as $V = CDC^T$, where C is an orthogonal matrix and D is a diagonal matrix with non-negative elements. Consider the transformation $y = Cx$. Then y is normally distributed with mean zero and covariance matrix D. The quadratic form $x^T A x$ transforms to $y^T F y$ where $F = C^T AC$. In terms of the new variables in y, which are independently distributed, the condition that $y^T F y$ has a χ^2-distribution is obviously $FDF = F$. Writing in terms of A and C, we have

$$C^T ACDC^T AC = C^T AVAC = C^T AC .$$

The last equality implies that $AVA = A$, which proves the desired result. The χ^2-distribution has degrees of freedom equal to the rank of VA.

Consider the particular quadratic form $x^T Vx$ where $V^- = CD^-C^T$ and D^- is obtained from D by replacing the non-zero elements by their reciprocals. Applying the test of Theorem 5.6, we find

$$V^-VV^- = CD^-C^TCDC^TCD^-C^T = CD^-C^T = V^- .$$

Hence, x^TV^-x has a χ^2-distribution with degrees of freedom equal to k, the rank of V.

Since $A^+ = A$ if A is idempotent and symmetric, and $A^+ = A^T$ if A is idempotent but not symmetric, no attempt is made to extend the theory of the distribution of quadratic forms of normal random vectors. An adequate and thorough expose' of this topic can be found in Graybill [48].

We now proceed to establish formulas for the conditional means and covariances which are valid even when the joint distribution is singular.

Theorem 5.7: Let $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be a partitioned zero mean normal random vector with

$$S = \text{cov} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} ,$$

cov $(x_1) = A$, and cov $(x_2) = C$, then the expected value of x_1, given that $x_2 = b$ and the covariance matrix of x_i, given that $x_2 = b$ are given by:

$$E(x_1 | x_2 = b) = BC^+ b$$

and

$$\text{cov } (x_1 | x_2 = b) = A - BC^+ B^T$$

Proof: We will derive the formulas for the conditional mean and covariance of x_1, given that $x_2 = b$, by representing x_1 in such a way that it is obvious what conditioning on x_2 means. We need only the rule for computing covariances under a linear transformation, i.e., if y has covariance matrix S, then My has covariance matrix MSM^T. Let $y = x_1 - BC^+ x_2$. Then the elements of the random vector y have zero means, and the covariance matrix of the composite vector

$$\begin{bmatrix} y \\ x_2 \end{bmatrix} = \begin{bmatrix} I & -BC^+ \\ 0 & I \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is

$$\begin{bmatrix} I & -BC^+ \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & 0 \\ -C^+ B^T & I \end{bmatrix} =$$

$$\begin{bmatrix} A - BC^+ B^T & B - BC^+ C \\ B^T - CC^+ B^T & C \end{bmatrix}$$

To establish that the off-diagonal blocks are 0, the general covariance matrix V is positive semidefinite. Hence, there exist a matrix P such that $V = P^TP$. Partitioning, we get

$$V = \begin{bmatrix} P_1^T \\ \\ P_2^T \end{bmatrix} (P_1 P_2) = \begin{bmatrix} P_1^T P_1 & P_1^T P_2 \\ \\ P_2^T P_1 & P_2^T P_2 \end{bmatrix}$$

and the column space of $P_2^TP_1$ lies in the column space of P_2^T which is the same as the column space of $P_2^TP_2$. Hence, without loss of generality we have that the columns of B^T lie in the column space of C. But, in that case $CC^+B^T = B^T$ since CC^+ is the projector of the column space of C. Hence, $B^T - CC^+B^T = 0$ implies that $B - BC^+C = 0$. Thus the covariance matrix becomes

$$\text{cov} \begin{bmatrix} y \\ x_2 \end{bmatrix} = \begin{bmatrix} A - BC^+B^T & 0 \\ 0 & C \end{bmatrix} .$$

Hence the covariance matrix of y is $A - BC^+B^T$, and y is independent of x_2. Because of this independence, it follows immediately that the conditional distribution of $x_1 = y + BC^+x_2$, given that $x_2 = b$, is normal with mean BC^+b and covariance that of y.

It should be noted that these formulas for conditional mean and covariance apply not only for the normal, but for any joint distribution for which zero correlation implies statistical independence.

5.3. Incidence Matrices:

In this section, the properties of the Penrose pseudo inverse of an arbitrary incidence matrix are examined in connection with the properties of the network flows in the corresponding directed graph and a simplified computational method for the pseudo inverse of an arbitrary incidence matrix is developed.

We begin by defining several terms:

By an incidence matrix, we shall mean a matrix which has exactly two nonzero entries that are 1 and -1 in each column of the matrix and has no zero rows.

Any two rows of an incidence matrix are said to be directly connected with each other if there is a column which has nonzero entries in both rows. Any two rows, i and j, of an incidence matrix are said to be indirectly connected with each other if there is a sequence of rows which starts with the i^{th} row and ends with the j^{th} row, $(i, k_1, k_2, \ldots, k_1, j)$, in which every two adjacent rows in the sequence are directly connected. Any two rows of an incidence matrix are said to be connected if they are directly or indirectly connected.

A connected component of an incidence matrix is a submatrix which consists of a set composed of rows, each pair of which are connected and none of which are connected with any other rows not in the set, and a set composed of all the columns which have nonzero entries in the rows in the set.

An incedence matrix is said to be a connected incidence matrix if it has only one connected component; otherwise it is said to be a separable incidence matrix. Then, by definition, a <u>separable incidence matrix</u> can be brought into the following form by suitable row and column interchanges;

$$T = \begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \cdots & \\ 0 & & & T_k \end{bmatrix},$$

where T_i, $i = 1, 2, \ldots, k$, is the matrix of the i^{th} connected component.

<u>Lemma 5.2</u>: If A is an $m \times n$ matrix of the form

$$A = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \cdots & \\ 0 & & & A_k \end{bmatrix},$$

where A_i, $i = 1, 2, \ldots, k$, is an $m_i \times n_i$ matrix and $\sum_{i=1}^{k} m_i = m$, $\sum_{i=1}^{k} n_i = n$, then the pseudo inverse of A is the $n \times m$ matrix given by

$$A^+ = \begin{bmatrix} A_1^+ & & & 0 \\ & A_2^+ & & \\ & & \cdots & \\ 0 & & & A_k^+ \end{bmatrix},$$

where A_i^+ is the pseudo inverse of A_i, $i = 1,], \ldots, 1$.

Proof: Let

$$\hat{A} = \begin{bmatrix} A_1^+ & & & 0 \\ & A_2^+ & & \\ & & \cdots & \\ 0 & & & A_k^+ \end{bmatrix},$$

Then we can easily verify that Penrose's four equations,

$$A\hat{A}A = A,$$
$$\hat{A}A\hat{A} = \hat{A},$$
$$A\hat{A} = (A\hat{A})^*,$$
$$\hat{A}A = (\hat{A}A)^*,$$

are all satisfied. Therefore, $\hat{A}$ satisfies all the conditions required for the pseudo inverse of A and by the uniqueness property of the pseudo inverse,

$$\hat{A} = A^+.$$

In the following discussion, we shall deal with only a connected incidence matrix, since the pseudo inverse of a separable incidence matrix can be derived by adjoining a set of the pseudo inverse of its connected components as shown in the above lemma and by making necessary row and column interchanges. This follows since for permutation matrices P_1 and P_2, $(P_1AP_2)^+ = P_1A^+P_2$.

Theorem 5.8: For any connected $m \times n$ incidence matrix T,

$$I - TT^{+} = \frac{1}{m} E, \tag{1}$$

where I is the $m \times m$ identity matrix and E is the $m \times m$ matrix whose elements are all equal to 1.

Proof: By the property of the pseudo inverse,

$$TT^{+}T = T \tag{2}$$

hence

$$(I - TT^{+}T) = 0 \tag{3}$$

This implies that if the k^{th} column of T has 1 in the i^{th} row and -1 in the j^{th} row, then the elements of $I - TT^{+}$ for columns i and j must be the same. Since T is connected, all columns of $I - TT^{+}$ are identical, and also since $I - TT^{+}$ is symmetrical by the property of the pseudo inverse (i.e., $TT^{+} = (TT^{+})^{*}$), all rows of $I - TT^{+}$ are identical. Hence, the elements in $I - TT^{+}$ are all identical. However, since $I - TT^{+}$ is idempotent, i.ec.,

$$(I - TT^{+})^{2} = I - TT^{+} - (TT^{+} - TT^{+}TT^{+}) = I - TT^{+}, \tag{4}$$

all elements in $I - TT^{+}$ are equal to $1/m$.

Lemma 5.3: Let T^{+} be the pseudo inverse of an $m \times n$ connected incidence matrix T and let e be the m-component column vector whose elements are equal to 1. Then

$$T^{+}e = 0 . \tag{5}$$

Proof: Since by definition

$$e^{*}T = 0 \tag{6}$$

the lemma follows from P28 in chapter 3.

Theorem 5.9: An $m \times n$ connected incidence matrix T contains at least one linearly independent set of $m - 1$ columns by which any column of T can be expressed uniquely as a linear combination of the columns in the set.

Such a set is called a basis of T.

Proof: Choose an arbitrary row, R_0, in T. Let R_1 be the set of all rows which are directly connected with the row R_0; R_1 be the nonempty set of all rows not in $R_0 \cup R_1 \cup \ldots \cup R_{i-1}$ which are directly connected with at least one of the rows in R_{i-1}. Since T is connected and $R_i\ R_j = \emptyset$ if $i \neq j$, every row in T belongs to one and only one of $R_0, R_1, \ldots, R_k$, $1 \leq k \leq m - 1$. Choose one column for every row in R_i which connects the row with any one of the rows in R_{i-1}, and let C_i be the set of such columns. Then the number of columns in C_i is equal to the number of rows in R_i and $C_i\ C_j = \emptyset$ if $i \neq j$, hence the set $C = C_1 \cup C_2 \cup \ldots \cup C_k$ consists of $m - 1$ columns.

Then every row is connected with the row R_0 by columns in C, hence every row is connected with every other row by columns in C.

Thus for any two rows in T, say the r_1^{th} row and the r_k^{th} row, there exists a sequence of columns in C, $(c_1, c_2, \ldots, c_{k-1})$, where $c_1 = \ell = 1, 2, \ldots, k-1$, directly connects the r_i^{th} row and the r_{1+1}^{th} row. (We shall assume that in the sequence of rows $(r_1, r_2, \ldots, r_k)$ no two columns are identical.) If the j^{th} column of T, denoted by c_j, has 1 in both the r_1^{th} row and -1 in the r_k^{th} row, then c_j is expressed as

$$c_j = \overset{k-1}{\underset{\ell=1}{\leftarrow}} (\pm c_\ell) , \tag{7}$$

where the plus sign is taken if c_ℓ has 1 in the r_ℓ^{th} row and the minus sign is taken if c_ℓ has -1 in the r_ℓ^{th} row. Such a sum with signs being adjusted according to the directions of arcs will be called a <u>sign-adjusted sum</u>.

Furthermore, every row in R_i has only one column in $C_1 \cup C_2 \cup \ldots \cup C_i$, $i = 1, 2, \ldots, k$, which has a nonzero element in the row. Hence, if a linear combination of columns in C is equal to the zero vector, the coefficients in the linear combination for the columns in C_k must be all equal to zero. This implies that the coefficients for the columns in C_{k-1} must also be all equal to zero, which in turn implies that the coefficients for the columns in C_{k-2} must also be all equal to zero, and so on. Thus, every coefficient must be equal to zero in order to have a linear combination of columns in C equal to the zero vector, hence C is linearly independent. Therefore, any column of T can be expressed uniquely as a linear combination of columns in C.

Using the above theorems, we derive a method of calculating T^+ as follows. Without loss of generality, let us assume that the first $m - 1$ columns of T form a basis and let T be partitioned into $[U: V]$, where U is an $m \times (n - m + 1)$ matrix whose columns are not in the basis. Also let T^+ be partitioned into $\begin{matrix}\hat{U}\\ \hat{V}\end{matrix}$, where $\hat{U}$ is an $(m - 1) \times m$ matrix which consists of the first $m - 1$ rows of T^+ and $\hat{V}$ is an $(n - m + 1) \times m$ matrix which consists of the remaining rows of T^+ in the sense that $\hat{U}$ is linearly independent and any row in $\hat{V}$ can be expressed uniquely as a linear combination of the rows in $\hat{U}$ (P28).

Let D be the $(m - 1 \times (n - m + 1)$ matrix such that

$$V = VD, \tag{8}$$

and let D^* be the transpose of D. Also let M be the $m \times m$ matrix which has $(m - 1)/ m$ for every diagonal element and $-1/m$ for every off-diagonal element. Then by Theorems 5.8 and (P28)

$$M = TT^+ = U\hat{U} + V\hat{V} = U\hat{U} + UDD^*\hat{U} = U(I + DD^*)\hat{U}. \tag{9}$$

This implies that for any j and j,

$$U_i(I + DD^*)\hat{U}_j = M_{ij}, \tag{10}$$

where U_i is the matrix U with the i^{th} row deleted, $\hat{U}_j$ is the matrix $\hat{U}$ with the j^{th} column deleted, and M_{ij} is the matrix M with the i^{th} row and the j^{th} column deleted, $i = 1, 2, . . ., m$; $j = 1, 2, . . ., m$.

Since the columns of the $m \times (m - 1)$ matrix U are linearly independent and any row of U can be expressed uniquely as the negative of the sum of the rest of the rows in U, the rank of the $(m - 1) \times (m - 1)$ matrix U_i is $m - 1$, i.e., U_i is nonsingular, for any $i = 1, 2, \ldots, m$. Also $I + DD^*$ is positive definite and hence nonsingular. Thus, the ordinary inverse of $U_i(I + DD^*)$ exists and U_j is uniquely determined by

$$U_j = [U_k(I + DD^*)]^{-1}M_{ij}. \tag{11}$$

The rest of the elements in T^+ can be derived as linear combinations of the elements in U_j by (P28) and Lemma 5.3. As an example, consider the following incidence matrix T:

$$T = \begin{bmatrix} 1 & 0 & 1 & -1 \\ -1 & 1 & 0 & 1 \\ 0 & -1 & -1 & 0 \end{bmatrix}$$

Here,

$$U = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} .$$

We arbitrarily set $i = 3$ and $j = 3$ for U_i and U_j. Then

$$U_3 = \left[\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \right\} \right]^{-1} \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 3/15 & -3/15 \\ 1/15 & 4/15 \end{bmatrix} .$$

By (P28) and lemma 5.3, we produce T^+ as follows:

$$T^+ = 1/15 \begin{bmatrix} 3 & -3 & 0 \\ 1 & 4 & -5 \\ 4 & 1 & -5 \\ -3 & 3 & 0 \end{bmatrix}$$

The following properties of the pseudo inverse of an incidence matrix may be derived from the above analysis.

First, we define the corresponding directed graph of a connected incidence matrix as a graph whose vertices and arcs have one-to-one correspondence with rows and columns, respectively, of the incidence matrix and each arc is directed from the i^{th} vertex toward the j^{th} vertex if the corresponding column of the matrix has 1 in the i^{th} row and -1 in the j^{th} row. The corresponding graph of the incidence matrix in the above example is shown in Fig. 1.

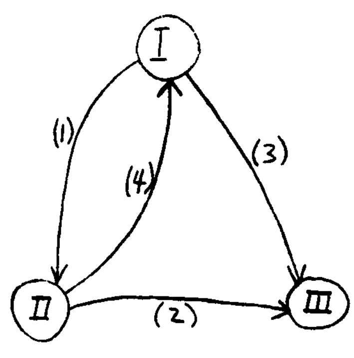

FIG. 1

If x is an n-component column vector which represents quantities or flows through n arcs in the graph, Tx represents the amount of the net inflow (or outflow if negative) to be made at each one of the m vertices. Since by Theorem 5.8,

$$TT^{+} = I - 1/m\, E, \tag{12}$$

where E is the $m \times m$ matrix whose elements are all equal to 1, the elements in the i^{th} column of T^{+} represent a set of quantities which flow through the n arcs when one unit of inflow is made at the i^{th} vertex and $1/m$ units of outflow is made at each one of the m vertices.

However, the quantities in the i^{th} column of T^{+} have an additional property. As brought out in the proof of Theorem 5.9, any column of T is expressible uniquely as the sign-adjusted sum of columns in a basis, and, by (P28), the corresponding row of T^{+} is also expressible uniquely by the same sign-adjusted sum of the corresponding rows in the basis of T^{+}. Hence, the flow quantity in the j^{th} arc is equal to the sign-adjusted sum of the flow quantities in a sequence of basis arcs (i.e., arcs whose corresponding columns are in the basis) which connect the same two vertices as the j^{th} arc does. This further implies that the sign-adjusted sum of the flow quantities in any sequence of arcs which connect a pair of vertices is identical for any given pair of vertices. This is equivalent to saying that the sign-adjusted sum of the flow quantities in the arcs in any loop is equal to zero, where a loop is a sequence of arcs which starts and ends with the same vertex, and every pair of adjacent arcs in the sequence have a common vertex.

Thus, the elements in the j^{th} row and the i^{th} column of T^{+} are equal to the quantity which flows through the j^{th} arc (in the direction of the arc) if the flows in the graph are made in such a way that the following two conditions are satisfied.

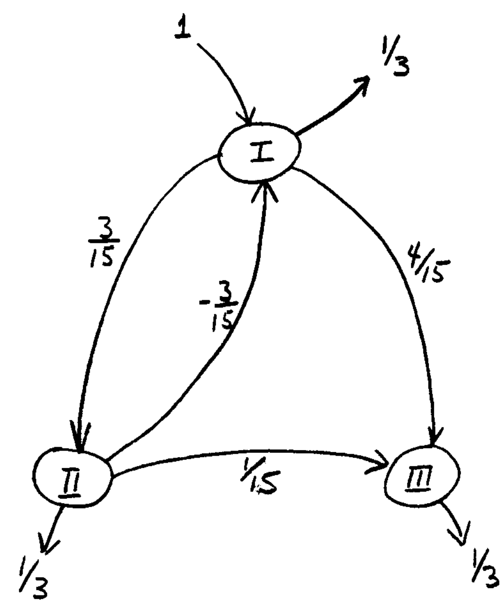

FIG. 2

Condition 1. One unit of inflow is made at the i^{th} vertex and $1/m$ units of outflow are made at each one of the m vertices.

Condition 2. For every pair of vertices the sign-adjusted sum of the flow quantities in a sequence of arcs which connect the two vertices is identical for any such sequences.

The two conditions uniquely determine the elements in T^{+} for any given directed graph in which the correspondence between rows of T^{+} and arcs of the graph and the correspondence between columns of T^{+} and vertices of the graph are fixed. To show this, let T be partitioned into $[U: V]$ as before, and let D be the matrix such that $V = UD$. Suppose that two $n \times m$ matrices $\hat{S}$ and $\overset{\downarrow}{S}$ both satisfy the above two conditions. Then, by Condition 1,

$$T(\hat{S} - \overset{\downarrow}{S}) = 0 . \tag{13}$$

Let the matrix $\hat{S} - \overset{\downarrow}{S}$ be partitioned into $\begin{matrix} \bar{U} \\ \bar{V} \end{matrix}$ where $\bar{U}$ is an $(m - 1 \times m$ matrix which corresponds to basis arcs and $\bar{V}$ is an $(n - m + 1) \times m$ matrix which corresponds to nonbasis arcs. Then, by Condition 2,

$$\bar{V} = D^{*}\bar{U}, \tag{14}$$

where D^{*} is the transpose of D. Hence,

$$T(\hat{S} - \overset{\downarrow}{S}) = U\bar{U} + UDD^{*}\bar{U} = U(I + DD^{*})\bar{U} = 0 \tag{15}$$

However, since the columns of U are linearly independent and the matrix $I + DD^{*}$ is nonsingular, as shown earlier, this implies that every element in $\bar{U}$ is zero. Hence, the matrix which satisfies the two conditions for a given graph is unique.

Fig. 2 is prepared from the first column of T^{+} in the above example.

The pseudo inverse of an arbitrary matrix possesses two types of least square properties, i.e., $\bar{x} = A^{+}y$ has the minimum norm, among all x's which minimize $||y - Ax||$. In our analysis of the pseudo inverse of a connected incidence matrix, this means that

(i) for any given y, $||y - Tx|| \geq ||y - TT^{+}y||$ for all x, and

(ii) among all x's which satisfy $||y - Tx|| = ||y - TT^{+}y||$, $x = T^{+}y$ has the minimum norm where $||x|| = (x^{*}x)^{1/2}$.

We shall show that the two conditions above can also be derived from these two types of least square properties of the pseudo inverses of connected incidence matrices. First, if (i) holds, Condition 1 must hold. For $e^{*}Tx = e^{*}TT^{+}y = 0$ for any x and y, and among all vectors $z = y - Tx$ whose elements add to the given constant $e^{*}y$, the vector whose elements are all equal to $(1/m)e^{*}y$ has the minimum norm, and by setting y equal to a unit vector Condition 1 follows.

If (ii) holds, Condition 2 must also hold. To show this, let T be partitioned into $[U:UD]$ as before. Let x be an n-component column vector of flow quantities and let it be partitioned into $\begin{matrix} x_1 \\ x_2 \end{matrix}$, where x_1 is an $(m - 1)$ -component column vector of flow quantities for basis arcs, and x_2 is an $(n - m + 1)$ -component column vector of flow quantities for nonbasis arcs. Since Condition 1 is satisfied, we must have for any given y,

$$Tx = Ux_1 + UDx_2 = y - \frac{e^{*}y}{m} e. \tag{16}$$

Since the columns of U are linearly independent, this implies that

$$x_1 + Dx_2 = z \tag{17}$$

where z is a given vector such that

$$Uz = y - \frac{e^{*}y}{m} e . \tag{18}$$

Let

$$L = x_1^* x_1 + x_2^* x_2 - \lambda^* (x_1 + Dx_2 - z), \tag{19}$$

where λ is an $(m - 1)$ -component column vector of Lagrange multipliers. Since $||x||$ is minimum, we must have

$$\frac{\partial L}{\partial x_1} = 2x_1 - \lambda = 0 \tag{20}$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - D^* \lambda = 0 , \tag{21}$$

hence,

$$x_2 = D^* x_1 ; \tag{22}$$

thus Condition 2 is satisfied.

The reader is referred to Berge [10] and Charnes and Cooper [20] for discussions on incidence matrices, graphs and their applications. See also Charnes, cooper, DeVoe and Learner [21] which is an interesting application of the pseudo inverse of an incidence matrix. The explicit form of the pseudo inverse of the distribution or transportation or dyadic matrix was first developed by A. Charnes, G. G. den Broeder, Jr., and R. E. Cline in 1956. (See, for example [Cline Ph.D. diss.]).

5.4. Stochastic Matrices

In this section we give an application of the Scroggs-Odell pseudo inverse to stochastic matrices where the spectral property inherited by the Scroggs-Odell pseudo inverse plays a very important role.

Let A be a stochastic matrix, i.e., $A \quad 0$ and $aj = j$, where $j = (1, 1, \ldots, 1)'$. A matrix A is said to be reducible if and only if there is a permutation matrix P such that

$$PAP^* = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$$

where B and D are square matrices. Otherwise the matrix A is called irreducible. For any reducible matrix there is a permutation matrix P such that

$$PAP^* = \begin{bmatrix} A_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & A_2 & & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & & A_k & 0 & \cdots & 0 \\ A_{k+1,1} & A_{k+1,2} & & A_{k+1,k} & A_{k+1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ A_{n1} & A_{n2} & \cdots & A_{nk} & A_{n,k+1} & \cdots & A_n \end{bmatrix}$$

where the A_i, $i = 1, 2, \ldots, n$ are irreducible. We say that A is completely reducible if and only if there is a permutation matrix P such that

$$PAP^* = \text{diag}\,(A_1, A_2, \ldots, A_n)$$

Theorem 5.10. Let A be a stochastic matrix. The necessary and sufficient conditions that A^+ be stochastic are that A be either completely reducible or irreducible and every non-zero eigenvalue of A lie on the unit circle.

Proof: We consider the necessity of the conditions. It is well known that the eigenvalues of a stochastic matrix lie in the closed unit disc. Consequently, it follows from Chapter 4, Theorem 5 that if A^+ is stochastic, then all non-zero eigenvalues of A (and A^+) must lie on the unit circle in the complex plane.

Let A be reducible. Then there exists a permutation matrix P such that

$$\hat{A} = PAP^*,$$

where $\hat{A}$ has the form (5.23). Since P is a permutation matrix $PP^* = I$. Thus $\hat{A}$ and A are similar and, hence, have the same eigenvalues. Due to the triangular form of $\hat{A}$, the eigenvalues of A are precisely those of all of the A_i, $i = 1, 2, \ldots, n$. Suppose that there is an i greater than k such that not all of the $A_{i1}, A_{i2}, \ldots, A_{i,i-1}$ are zero. But in this case, the spectral radius of A_i is less than the spectral radius of A. Thus the eigenvalues of A_i are in modulus less than 1. This is a contradiction. Hence A is completely reducible. Thus, the proof of the necessity is concluded.

The following lemma will be needed in the proof of the sufficiency.

Lemma 5.4: If A is stochastic, irreducible and has all of its non-zero eigenvalues on the unit circle, then the elementary divisor of A corresponding to zero is at most of first degree.

Proof of the Lemma. It follows from a well known redult due to Frobenius, that the minimum equation for A is of the form $x^p (x^h - 1) = 0$. Factoring the polynomial in this equation,

$$x^p (x^h - 1) = (x - 1)(x^{p+h-1} + x^{p+h-2} + \ldots x^p) \qquad (1)$$

Thus

$$(A - I)(A^{p+h-1} + A^{p+h-2} + \ldots + A^p = 0 \qquad (2)$$

Now a Jordan form for A is

$$C = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \omega_1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \omega_2 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & \omega_{h-1} & 0 \\ 0 & 0 & 0 & \ldots & 0 & N \end{bmatrix}$$

where ω_i, $i = 1, 2, \ldots, h - 1$, are the h-th roots of unity different from 1 and N is a p by p matrix whose elements are all zero except for the diagonal above the principal diagonal. The first $p - 1$ elements in the diagonal above the principal diagonal are 1's and the other elements are zero. Then

$$A^{p+h-1} + A^{p+h-2} + \ldots + A^p = P(C^{p+h-2} + \ldots + C^p)\, P^{-1}$$

$$- P\,(\mathrm{diag}\,(h, 0, 0, \ldots 0))\, P^{-1} \qquad (3)$$

Thus the i-th row of the sum of the matrices on the left in (3) is $hp_{i1}P_1^{-1}$, where P_1^{-1} is the first row of P^{-1}. But each matrix in the sum on the left in (3) is a stochastic matrix. Hence the sum of the elements in a given row of the sum is h. Now the first column of P is an eigenvector of A corresponding to 1. Hence we may take $p_{i1} = 1$, for $i = 1, 2, \ldots, n$. Thus the sum of the elements of the first row of P^{-1} is 1. Now if $p > 1$, consider the sum

$$A^{p+h-2} + A^{p+h-3} + \ldots + A^{p-1} = PDP^{-1}, \tag{4}$$

where $d_{11} = h$, $d_{h+1,h+p} = 1$ and all other elements of D are zero. Then the i-th row of PDP^{-1} is $hp_{i1}P_1^{-1} + P_{i,h+1}P_{h+p}^{-1}$. The summands on the left in (4) are each stochastic. Thus the sum of each row of PDP^{-1} must be h. But $p_{i,h+1}$ is not zero for at least one value of i between 1 and n and the row P_{h+p}^{-1} is not identically zero. Thus we have contradicted the fact that the sum of the elements of P_1^{-I} is 1. Hence $p \leq 1$.

Returning to the proof of the sufficiency portion of Theorem 5.10, suppose that A is completely reducible and all of the non-zero eigenvalues of A lie on the unit circle. Then there is a permutation matrix P such that $PAP^* = \text{diag}(A_1, A_2, \ldots, A_g)$. Since each of the A_i is irreducible and the nonzero eigenvalues of each of the A_i lie on the unit circle, the non-zero eigen values of A_i are precisely the h_i-th roots of unity. Thus it follows that the minimum polynomial for A is of the form $x^p(x^k - 1)$, where i = l.c.m.

$(h_1, h_2, \ldots, h_g)$. From the lemma, either $p = 0$ or $p = 1$. If $p = 0$, then A is non-singular and $A^+ = A^{-1} = A^{k-1}$. If $p = 1$, then A is singular and $A^+ = A^{k-1}$. In either case, A^+ is stochastic.

5.5. A Generalization of the Gauss-Markov Theorem

Consider the linear model

$$y = Hx + e$$

where y is a real p by 1 vector of observations, x is a real unknown n by 1 state vector, e is a real p by 1 random error vector, and H is a p by n known real matrix. Also $E(E) = \phi$ and $E(ee^T) = V$ where E denotes the expected value operation, ϕ denotes the null matrix (or vector), and V is a known real symmetric positive definite matrix.

We seek a linear, minimum variance, unbiased estimate $\hat{x}$ of x. That is, we are to find a matrix B such that $\hat{x} = By$, $E(\hat{x}) = x$, and $V = E[(\hat{x} - x)\ (\hat{x} - x)^T]$ is minimum in the sense that if z is any linear unbiased estimate of x, then $q^T[V_z - V_{\hat{x}}]\ q > 0$ for any p by 1 vector $q \neq 0$. V_z is the corresponding covariance matrix of z, which is a real symmetric positive definite matrix. These conditions imply that $E(\hat{x}) = BHx = x$ so that $BH = I$, where I is the $n \times n$ identity.

If the rank of H is $p < n$, we cannot require that $E(\hat{x}) = x$, since, in this case H has no left inverse. We can, however, modify this requirement by requiring that the norm $||E(\hat{x}) - x||$ be minimum. The properties remain unchanged for complex matrices if we replace transpose by conjugate transpose.

To facilitate reading, we list some properties of the Penrose pseudoinverse used in obtaining this result.

P1) For every matrix A there exists a unique matrix A^+ such that $AA^+A = A$

$$A^+AA^+ = A^+$$

$$(A^+A)^T = A^+A$$

$$(AA^+)^T = AA^+.$$

We call A^+ the Penrose pseudo-inverse of A.

P2) $(AC)^+ = C_1^+ A_1^+$ where $AC = A_1C_1$, $C_1 = A^+AC$, and $A_1 = AC_1C_1^+$

P3) $(A^+)^T = (A^T)^+$

P4) All solutions of the matrix equation $AXB = C$ are given by $x = A^+CB^+ + Y - A^+A\,y\,BB^+$ if and only if $AA^+CB^+B = C$ where Y has the dimension of X.

P5) Range of A^T equals the range of A^+, that is $R(A^T) = R(A^+)$. A^+A and AA^+ are, respectively, the projection operators on the range spaces of A^+ and A.

P6) For any $n \times n$ matrix A and vector, z, $z = z_1 + z_2$ $z_1 \epsilon R(A^+)$, $z_2 \epsilon N(A)$, and x_1 is orthogonal to z_2.

We are now ready to establish a generalization of the Gauss-Markov theorem.

Theorem 5.11: Consider the linear model described by the vector equation

$$\underset{p\times 1}{y} = \underset{p\times m}{H}\;\underset{m\times 1}{x} + \underset{p\times 1}{e}$$

where $E(e) = \phi$ and $E(ee^T) = V$ is positive definite. The minimum mean-square-error linear estimate $\hat{x}$ of x is given by:

$$\hat{x} = M^+ H^T V^{-1} y$$

with

$$V_{\hat{x}} = M^+$$

where

$$M = H^T V^{-1} H .$$

<u>Proof</u>: We require that $\hat{x} = By$ and $E(\hat{x}) = x$ whenever $x \epsilon R(H^T)$. These requirements imply that $E(\hat{x}) = BHx$ and (P5) implies that for x in $R(H^T)$,

$$H^+ Hx = BHx = x.$$

Let $x = x_1 + x_2$ where $x_1 \epsilon R(H^T)$, $x_2 \epsilon N(H)$. Then

$$||E(\hat{x}) - x|| = ||BHx_1 + BHx_2 - x|| = ||BHx_2 - x_2|| = ||x_2||$$

It follows that $||E(\hat{x}) - x||$ is minimum for $x \not\epsilon R(H^T)$. The covariance matrix $V_{\hat{x}}$ of the estimate $\hat{x}$ is given by $V_{\hat{x}} = BVB^T$ and must be minimized subject to the constraint $GH = H^+H$. To do this we adjoin the constraint $BH = H^+H$ to BVB^T using a matrix Lagrange multiplier Λ and find conditions necessary to minimize

$$Q = BVB^T + \Lambda^T [H^+H - H^T B^T] + [H^+H - BH] \Lambda.$$

Employing the variational technique [38] we obtain the first variation δQ,

$$\delta Q = \delta B [VB^T - H\lambda] + [BV - \lambda^T H^T] \delta B^T.$$

Since δB is arbitrary, we find that setting $\delta Q = \phi$ implies

$$BV - \lambda^T H^T = \phi$$

or

$$B = \lambda^T H^T V^{-1}.$$

Multiplying the latter by H we obtain

$$H^+ H = \lambda^T H^T V^{-1} H$$

so that using (P3) and setting $H^T V^{-1} H = M$ we have

$$\lambda^T = H^+ H M^+ + y\,[I - MM^+] = M^+ + y\,(I - MM^+)$$

where y is arbitrary to within having the dimension of λ^T. Assume that the rank of H is $q \leq \min(n, p)$. Then

$$B = \lambda^T H^T V^{-1}$$

$$= \{M^+ + y\,[I - MM^+]\}\; H^T V^{-1}$$

We need to establish a workable form for M^+. To do this apply (P1) with $A = H^T V^{-1}$ and $C = H$. We get

$$C_1 = (H^T V^{-1})^+ \, H^T V^{-1} H$$

$$A_1 = H^T V^{-1} \, (H^T V^{-1} \, (H^T V^{-1})^+ \, H^T V^{-1} H \, [(H^T V^{-1})^+ \, H^T V^{-1} H]^+$$

$$= H^T V^{-1} H \, [(H^T V^{-1})^+ \, H^T V^{-1} H]^+$$

Hence

$$M^+ = [(H^T V^{-1})^+ \, H^T V^{-1} H]^+ \, [H^T V^{-1} H \, \{(H^T V^{-1})^+ \, H^T V^{-1} H\}^+]^+$$

Therefore

$$B = \{M^{+} + y\,[I - MM^{+}]\}\,H^{T}V^{-1}$$

$$= M^{+}H^{T}V^{-1} + y\,[I - H^{T}V^{-1}H\,[(H^{T}V^{-1})^{+}\,H^{T}V^{-1}H]^{+}$$

$$H^{T}V^{-1}H\,\{(H^{T}V^{-1})^{+}\,H^{T}V^{-1}H\}^{+}]^{+}\,H^{T}V^{-1}.$$

To establish the second term is ϕ, i.e. $[I - MM^{+}]\,H^{T}V^{-1} = \phi$. We need $MM^{+}\,H^{T}V^{-1} = H^{T}V^{-1}$. Since $[I - MM^{+}]$ is an orthogonal projection on the null space of M^{+}, we need to show that $N(M^{+}) = N(H)$. Since $M = H^{T}V^{-1}H$, then it certainly follows that $n(M) = N(M^{T})$. Also note that $N(M) = N(H)$. Thus suppose there exists an $x \epsilon N(M)$ such that $x \notin N(H)$. Since V^{-1} is positive definite, V^{-1} does not rotate Hx into the null space of H^{T}. Hence $H^{T}V^{-1}\,Hx \neq 0$ which implies $x \notin N(M)$. This is a contradiction. Thus $N(M) = N(H)$. Now $N(M) = N(M^{T}) = N(M^{+})$ which implies $N(M^{+}) = N(H)$ and consequently $(I - MM^{+})\,H^{T}V^{-1} = \phi$ since $R(H^{T}) = N(H)^{\perp}$. Hence

$$\hat{x} = By = M^{+}H^{T}V^{-1}y$$

with covariance matrix

$$V_{\hat{x}} = BVB^{T} = M^{+}H^{T}V^{-1}HM^{+T}$$

$$= M^{+}MM^{+T} = M^{+}.$$

There are two special cases where the formulas for $\hat{x}$ and $V_{\hat{x}}$ reduce very nicely.

Case 1: Rank $H = n \leq p$. In this case $H^{+}H = I$. Thus

$$\hat{x} = M^{+}H^{T}V^{-1}y$$

$$= M^{-1}H^{T}V^{-1}y$$

$$= (H^{T}V^{-1}H)^{-1} H^{T}V^{-1}y .$$

and

$$V_{\hat{x}} = (H^{T}V^{-1}H)^{-1} H^{T}V^{-1}H(H^{T}V^{-1}H)^{-1}$$

$$= (H^{T}V^{-1}H)^{-1} .$$

Case 2: Rank $H = p \leq n$. In this case $HH^{+} = I$ and substituting into the four defining equations establishes that

$$(H^{T}V^{-1})^{+} = VH^{T+}$$

Thus

$$M^{+} = [VH^{T+}H^{T}V^{-1}H]^{+} [H^{T}V^{-1}H \{VH^{T+}H^{T}V^{-1}H\}^{+}]^{+}$$

$$= H^{+} (H^{T}V^{-1}HH^{+})^{+}$$

$$= H^{+} (H^{T}V^{-1})^{+}$$

$$= H^{+}VH^{T+} .$$

Hence

$$\hat{x} = M^{+}H^{T}V^{-1}y$$

$$= H^{+}VH^{T+}H^{T}V^{-1}y$$

$$= H^{+}y ,$$

and

$$V_{\hat{x}} = H^{+}VH^{+T} .$$

It is of interest to compare the least squares estimate of the state vector to that of the minimum variance estimate of the state vector. Magness and McGuire [63] have been able to give an extensive anlaysis in comparing these two estimates whenever the regression matrix of the linear model is of full-rank (columns linearly independent). They were able to establish the inequality

$$V_{LS} \leq \frac{1}{4} (\lambda_{max} + \lambda_{min}) \left(\frac{1}{\lambda max} + \frac{1}{\lambda min}\right) V_{MV}$$

where V_{LS} and V_{MV} are the covariance matrices of the least squares estimate and minimum variance estimate, respectively. λ_{max} and λ_{min} are the maximum and minimum eigenvalues of the correlation matrix ρ of the error vector. The above inequality places an upper bound on how much is lost by use of the least squares estimate of the state vector to that of the minimum variance estimate of the state vector.

In the following theorem it will be shown that the least-squares estimate of the state vector will have the same covariance matrix as that of the mean-square-error estimate of the state vector, whenever the regression matrix of the linear model has all of its rows linearly independent.

<u>Theorem 5.12</u>: Consider the linear model described by the vector equation

$$\underset{px1}{y} = \underset{pxn}{H} \; \underset{nx1}{x} + \underset{px1}{e}$$

where $E(e) = \phi$, $E(ee^T) = V$ is positive definite, and $R(H) = p$. Then the covariance matrix of the least-squares estimate of the state vector equals the covariance matrix of the mean-square-error estimate of the state vector.

Proof: The least squares estimate of the state vector is

$$\hat{x}_{LS} = (H^T H)^+ H^T Y$$

$$= H^+ Y.$$

The corresponding covariance matrix is

$$V_{LS} = H^+ V H^{T+}.$$

The mean-square-error estimate of the state vector is by Theorem 5.11, Case 2,

$$\hat{x} = H^+ y$$

The corresponding covariance matrix is

$$V = H^+ V H^{T+}.$$

Thus it can be seen there is no loss in using the least squares estimate whenever the rows of the regression matrix are linearly independent.

5.6 An Application of a Pseudoinverse to Testing Hypothesis

Consider the matrix A which is $n \times p$. Let A^- be a $p \times n$ matrix satisfying the following equations:

$$AA^-A = A$$

$$A^-AA^- = A^-$$

$$(AA^-)^T = AA^-.$$

A^- will be called a pseudo inverse of A. The above equations imply that the null space of A^- is the orthogonal complement of the range of A, and AA^- is an orthogonal projection operator on the range of A. We will use a matrix with the above properties to establish a general method for testing a hypothesis about a linear model.

We shall consider the linear model

$$y = Hx + V$$

where y is a real $n \times 1$ vector of observations, x is a real unknown $r \times 1$ state vector, V is a real $n \times 1$ random error vector, and H is an $n \times r$ known real matrix. In addition V is distributed $N(0, \sigma^2 I)$.

Let Ω be the p-dimensional vector space spanned by the columns of H. Assume $V \epsilon \Omega$, and that ω is a q-dimensional subspace of Ω spanned by the columns of H_1.

We wish to test the following hypothesis:

$$H_0\colon\ E(y) \ \epsilon\ \omega \ ; \ \text{i.e.,}\ H_1\alpha = \eta$$

We define the likelihood-ratio statistic λ to be:

$$\lambda = \max_{\Omega} p(y) \;/\; \max_{\omega} p(y), \text{ where } p(y) \text{ is}$$

the probability density function of y. The likelihood-ratio test consists in rejecting H_0 if $\lambda > \lambda_2$, where λ_2 was chosen such that the $P_r(\lambda > \lambda_2) \leq \alpha$. Since y is distributed $N(Hx, \sigma^2 I)$ it has been shown that

$$\max_{\Omega} P(y) = (2\pi || y - \hat{u}||)^{-n/2} \; e - \frac{1}{2n}$$

$$\max_{\omega} P(y) = (2\pi || y - \hat{\eta} \quad)^{-n/2} \; e - \frac{1}{2n}$$

where

$$\hat{u} = H\hat{x} \text{ and } \hat{x} \text{ satisfies } H^T H\hat{x} = H^T y \tag{1}$$

$$\hat{\eta} = H_1\hat{\alpha} \text{ and } \hat{\alpha} \text{ satisfies } H_1{}^T H_1\hat{\alpha} = H_1{}^T y \tag{2}$$

But, since $H^T HH^- = H^T$ and $N(A^-) = R(A)^{\perp}$ it follows that equations (1) and (2) have a solution given by

$$\hat{x} = H^- y$$

$$\hat{\alpha} = H^-{}_1 y.$$

Hence, λ can be written as

$$\lambda = \left[\frac{||y - H_1\alpha||^2}{||y - H\hat{x}||^2} \right]^{n/2}$$

We now define the statistic F to be

$$F(\lambda) \quad = \quad \frac{n-p}{p-q} \; (\lambda^{2/n} \; 0 \; 1).$$

Since $F(\lambda)$ is a single-valued everywhere increasing function of λ, the λ-test is equivalent to rejecting H_0 if and only if $F > F(\lambda_0)$, where $F(\lambda_0)$ is chosen such that $P_r\{F > F(\lambda_0)\} \leq \alpha$. We will now show that F is the central F if and only if H_0 is true. Now

$$F \quad = \quad \frac{n-p}{p-q} \; \left[\frac{||y - H_1\hat{\alpha}||^2}{||y - H\,\hat{x}||^2} - 1\right]$$

or

$$F \quad = \quad \frac{n-p}{p-q} \; \left[\frac{||y - H_1\hat{\alpha}||^2 - ||y - H\hat{x}||^2}{||y - H\hat{x}||^2}\right]$$

We will rewrite F using the fact that $\hat{x} = H^{-}y$, $\hat{\alpha} = H^{-}_{1}y$, $(HH^{-})^T = HH^{-}$ which gives

$$||y - H_1\hat{\alpha}||^2 \quad = \quad (y - H_1\hat{\alpha})^T \; (y - H_1\hat{\alpha}).$$

We have

$$\begin{aligned} ||y - H_1\hat{\alpha}||^2 \quad &= \quad (y - H_1H_1^{-}\,y)^T \; (y - H_1H_1^{-}\,y) \\ &= \quad y^Ty - y^TH_1^{-T}H_1^{\,T}y - y^TH_1H_1^{-}y + y^TH_1^{-T}H_1^TH_1H_1^{-}y \\ &= \quad y^Ty - y^TH_1H_1^{-}y \\ &= \quad y^T[I - H_1H_1^{-}]y. \end{aligned}$$

Likewise

$$||y - H\hat{x}||^2 = y^T [I - HH^-]y.$$

Therefore

$$||y - H_1\hat{\alpha}||^2 - ||y - H\hat{x}||^2 = y^T [HH^- - H_1H_1^-]y,$$

so that

$$F = \frac{n-p}{p-q} \left[\frac{y^T[HH^- - H_1H_1^-]y}{y^T[I - HH^-]y} \right]$$

We will now show that $HH^- - H_1H^-$ and $I - HH^-$ are symmetric and idempotent. To do this we use the facts that $(HH^-)^T = HH^-$ and $HH^-HH^- = HH^-$.

$$(HH^- - H_1H_1^-)^T = HH^- - H_1H_1^-$$

$$(I - HH^-)^T = I - HH^-$$

$$(HH^- - H_1H_1^-)(HH^- - H_1H_1^-) = HH^- - H_1H_1^-HH^- - HH^-H_1H_1^- + H_1H_1^- .$$

We now show that $HH^-H_1H_1^- = H_1H_1^-HH^- = H_1H_1^-$. To do this let $Z = Z_1 + Z_2 + Z_3$ where

$Z_1 \varepsilon \Omega^\perp$

$Z_2 \varepsilon \omega$

$Z_3 \varepsilon \omega^\perp$ with respect to Ω

$$(H_1H_1^-HH^-) Z = H_1H_1^- (Z_2 + Z_3) = Z_2$$

$$(HH^-H_1H_1^-) Z = H^-Z_2 = Z_2$$

Now

$$G_k G_k^- Z_2 = Z_2.$$

Thus

$$HH^- H_1 H_1^- + H_1 H_1^- HH^- = 2H_1 H_1^- .$$

Hence

$$(HH^- - H_1 H_1^-)(HH^- - H_1 H_1^-) = HH^- - H_1 H_1^- .$$

Also

$$[I - HH^-][I - HH^-] = I - HH^- - HH^- + HH^- HH^- = I - HH^- ,$$

and

$$[HH^- - H_1 H_1^-][I - HH^-] = HH^- - H_1 H_1^- - HH^- + H_1 H_1^- HH^- = 0 .$$

Hence, we have that $y^T[I - HH^-]y$ and $y^T[HH^- - H_1 H_1^-]y$ are independently distributed as $X^2(n - p, \delta_1)$ and $X^2 (p - q, \delta_2)$, respectively, where

$$\begin{aligned} \delta_1 &= [E(y)^T [I - HH^-] E(y)] = x^T H^T [I - HH^-] Hx \\ &= x^T [H^T H - H^T (HH^- H)] x \\ &= x^T [H^T H - H^T H] x = 0, \end{aligned}$$

and

$$\delta_2 = E(y)^T [HH^- - H_1 H_1^-] E(y) .$$

Hence

$$\frac{n-p}{p-q}\ \frac{y^T[HH^- - H_1H_1^-]y}{y^T\ [I - HH^-]\ y}$$

is distributed as $F(p - q, n - p, \delta_2)$.

Now if H_0 is true, $E(y) = H_1\alpha = \eta$ so that

$$\delta_2 = \eta^T\ [HH^- - H_1H_1^-]\ \eta.$$

Since $\eta\varepsilon\omega$, $HH^-\eta = \eta$ and $H_1H_1^-\eta = \eta$, therefore $\delta_2 = \eta^T[\eta\ \eta] = 0$. Suppose $\delta_2 = 0$. Let $E(y) = Z$, where $Z = z_1 + Z_2$, $Z_1\varepsilon\omega$, $Z_2\varepsilon\omega$. This implies that

$$(Z_1^T + Z_2^T)\ (HH^- - H_1H_1^-\ (Z_1 + Z_2) = 0$$

$$(Z_1^T + Z_2^T)\ (Z_1 + Z_2 - Z_1) = 0$$

$$(Z_1^T + Z_2^T)\ Z_2 = 0$$

$$Z_1^TZ_2 + Z_2^TZ_2 = 0.$$

But $Z_1^TZ_2 = 0$, since Z_1 is in the orthogonal complement of Z_2. Hence, $Z_2^TZ_2 = 0$, so that $Z_2 = 0$. We conclude that if $\delta_2 = 0$, $E(y)\varepsilon\omega$. We have thus established that H_0 is true if and only if $\delta_2 = 0$. Hence F is the central F if and only if H_0 is true.

It should be noted that the above test could be done with the Penrose pseudoinverse. The more general pseudo inverse A^- was given to indicate the possibility of defining a pseudo inverse and adapting it to a particular situation.

5.7 Application To Estimable Functions

Consider the linear model $Y - HX + V$ where Y is an n by 1 real vector of observations, H is an n by p known real matrix of rank $q \leq \min(n, p)$, X is a p by 1 state vector, and V is a real n by 1 error vector. Also let $E(V) = \phi$ and $E(VV^T) = \alpha^2 I$. Suppose it is desired to estimate the state vector by the method of least-squares. Thus it becomes necessary to minimize $V^T V = (Y - HX)^T (Y - HX)$, which gives the normal equations $(H^T H)X = H^T Y$. A simple argument can be used to show that this system is consistent and thus the general solution is $\hat{X} = H^+ Y + (I - H^+ H)Z$ where Z is arbitrary. This general solution implies that there are infinitely many solutions. To the statistician this is undesirable for two researchers with the same data, both using the same method of estimation, can draw different conclusions. Also it can be seen from observing the general solution that no unbiased estimate of X exists unless H is of rank p which is also undesirable.

It would seem natural to investigate whether there exists an unbiased estimate of any linear combinations of the elements of X. Before proceeding further we shall formulate two useful definitions.

Definition 5.1: A parameter is said to be estimable if there exists an unbiased estimate of the parameter.

Definition 5.2: A parameter is said to be linearly estimable if there exists a linear combination of the observations whose expected value is equal to the parameter.

Let A be a matrix such that $A(I - H^+H) = \phi$, then $H^+HA^T = A^T$. Hence, the columns of A^T belong to $R(H^+)$ and thus $A\hat{X} = AH^+Y + A(I - H^+H)Z = AH^+Y$ which implies $E(A\hat{X}) = E(AH^+Y) = AH^+HX = AX$. Hence, for any A such that $A(I - H^+H) = \phi$, the parameter AX is an estimable function.

<u>Theorem 5.13</u>: Let H be n by p of rank $q \leq \min(n, p)$, then AX is estimable if and only if there exists a solution for r in the equations

$$H^T Hr = a_i^T$$

where

$$A = (a_1^T, a_2^T, \ldots, a_n^T)^T.$$

<u>Proof</u>: Partition A such that each a_i is a 1 by p row vector. Suppose AX is estimable, then there exists a B such that $E(BY) = AX$ which implies $BHX = AX$ for every X. Hence it follows $BH = A$ which implies $H^T B^T = A^T$. Thus the columns of A^T belong to the column space of H^T and consequently $H^+Ha_i^T = a_i^T$. The estimate of a_iX is a_iH^+Y. It is unbiased since

$$E(a_iH^+Y) = a_iH^+HX = a_iX.$$

Now $a_i^T \varepsilon R(H^+)$, but $R(H^+) = R(H^T)$, hence, there exists a vector z such that

$$H^T z = a_i^T .$$

Hence, the rank of the matrix H^T is equal to the rank of the augmented matrix $[H^T|a_i^T]$. Thus the rank of H^TH equals the rank of $[H^TH|a_i^T]$ which implies $H^THr = a_i^T$ has a solution. Conversely, if $H^THr = a_i^T$ has a solution, we let $b = Hr$. Then

$$E(b^TY) = E(r^TH^TY) = r^TH^THX = a_iX = E(a_iH^+Y).$$

Hence, a_i is a row of A if and only if there exist a solution for r in the equations $H^THr = a_i^T$, $i = 1, 2, \ldots, n$.

Theorem 5.14: Let H be n by p of rank $q \leq \min(n, p)$, then the best linear unbiased estimate for any estimable function AX is AH^+Y.

Proof: Assume that the best linear unbiased estimate of AX is $CY = (AH^+ + B)Y$. Now CY is completely general since B is general. We must determine the matrix B such that

$$E(CY) = E\,[(AH^+ + B)Y] = AH^+HX + BHX$$

$$= AX + BHX = AX.$$

Hence $BH = \phi$. To show that $B = \phi$ we must minimize the variance of CY.

$$\mathrm{Cov}(CY) = E[(CY - AX)\ (CY - AX)^T]$$

$$= E[CYY^TC^T = AXY^TC^T - CYX^TA^T + AXX^TA^T]$$

$$= C(\sigma^2I + HXX^TH^T)\ C^T - AXX^TH^TC^T - CHXX^TA^T + AXX^TA^T$$

$$= (AH^+ + B)\ (\sigma^2I + HXX^TH^T)\ (AH^+ + B)^T - AXX^TH^T\ (AH^+ + B)^T$$

$$- (AH^+ + B)\ HXX^TA^T + AXX^TA^T$$

$$= \sigma^2 \ (AH^+ + B) \ (AH^+ + B)^T$$

$$= \sigma^2 \ (AH^+ \ H^{+T}A^T + BH^{+T} + AH^+B^T + BB^T).$$

Now $BH = \phi$ implies that $H^TB^T = \phi$ so that the columns of B^T are in the null space of H^T which is the same as the null space of H^+. Hence it follows that $H^+B^T = \phi$. Thus

$$\text{Cov (CY)} \ = \ AH^+H^{+T}A^T + BB^T.$$

Hence to minimize var (CY) we must minimize the diagonal elements of BB^T. But, they are all non-negative, hence to minimize the var(CY), we must take $B = \phi$. Thus $C = AH^+$ and AH^+Y is the best linear unbiased estimate of AX.

Definition 5.3: The estimable functions a_iX, $i = 1, 2, \ldots, k$ are said to be linearly independent estimable functions if the a_i are linearly independent.

Theorem 5.15: Let H be an n by p matrix of rank $q \leq \min (n, p)$, then there are exactly q linearly independent estimable functions.

Proof: a_iX is estimable if and only if $a_i^T \in R(H^T)$, which implies that $a_i[I - H^+H] = 0$. Also, by Theorem 3.1, a_iX is estimable if and only if there exist a solution for r in the equation $H^THr = a_i^T$. Let $r_1, r_2, \ldots, t_t$ be such that $H^THr_1 = a_1^T, \ldots, H^THr_t = a_t^T$. Then

$$H^TH \ (r_1, r_2, \ldots, r_t) = (a_1^T, a_2^T, \ldots, a_t^T).$$

But H^TH is of rank q which implies that $(a_1^T, a_2^T, \ldots, a_t^T)$ is at most of rank q. But each $a_i^T \in R(H^T)$ so that each a_iX is estimable. But the rank of H^+ is q which implies there are q linearly independent a_i^T's. Hence, there are exactly q linearly independent estimable functions.

Theorem 5.16: Let H be an n by p matrix of rank $q \leq \min(n, p)$. Let AX be an estimable function where A is k by p of rank q. If BX is an estimable function, then the rows of B are linear bombination of the rows of A.

Proof: This follows immediately from Theorem 5.15.

It is interesting to note that since the rows of H are elements of $R(H^T)$ which implies the rows of H are elements of $R(H^T)$, that HX is an estimable function. Also since BH is contained in the row space of H, then BHX is an estimable function. In fact $A_1A_2 \ldots A_nHX$ is an estimable function. Furthermore, it is obvious that the best linear unbiased estimate of a linear combination of estimable functions is given by the same linear combination of the best linear unbiased estimates of the estimable functions.

Variance and Covariance of Estimable Functions

Theorem 5.17: If A_1X and A_2X are two estimable functions, the respective covariances of the best linear unbiased estimates are $\sigma^2A_1H^+H^{+T}A_1^T$ and $\sigma^2A_2H^+H^{+T}A_2^T$. The covariance of the estimates of A_1X and A_2X is equal $\sigma^2A_1H^+H^{+T}A_2^T$.

Proof: From the proof of Theorem 5.14 we see that the covariance of the estimate of A_1X is $\sigma^2 A_1 H^+ H^{+T} A_1^T$. Similarly the covariance of the estimate of A_2X is $\sigma^2 A_2 H^+ H^{+T} A_2^T$. The covariance of the estimates of A_1X and A_2X is given by:

$$\begin{aligned}
&\mathrm{Cov}\ [A_1H^+Y,\ A_2H^+Y] \\
&= E\ [(A_1H^+Y - A_1X)\ (A_2H^+Y - A_2X)^T] \\
&= E\ [A_1H^+YY^TH^{+T}A_2^T - A_1XY^TX^{+T}A_2 - A_1H^+YX^TA_2^T + A_1XX^TA_2^T] \\
&= A_1H^+\ (\sigma^2 I + HXX^TH^T)H^{+T}A_2^T - A_1XX^TH^TH^{+T}A_2^T - A_1H^+HXX^TA_2^T + \\
&\qquad A_1XX^TA_2^T \\
&= \sigma^2 A_1H^+H^{+T}A_2^T + A_1XX^TA_2^T - A_1XX^TA_2^T - A_1XX^TA_2^T + A_1XX^TA_2^T \\
&= \sigma^2 A_1H^+H^{+T}A_2^T\ .
\end{aligned}$$

Theorem 5.18: Let $Y = HX + V$, where H is an n by p matrix of rank $q \le \min(n, p)$ and V is distributed $N_n(0, \sigma^2 I)$, then the quantity

$$(n - q)\ \hat{\sigma}^2 = (Y - H\hat{X})^T\ (Y - H\hat{X})$$

is distributed as a chi-square variate with $n - q$ degrees of freedom. In symbols,

$$\chi^2\ (n - q,\ \lambda = 0).$$

Proof:

$$(Y - H\hat{X})^T (Y - H\hat{X}) = (Y - HH^+Y)^T (Y - HH^+Y)$$

$$= Y^T (I - HH^+) (I - HH^+)Y = Y^T (I - HH^+)Y.$$

But $(I - HH^+)$ is idempotent and hence $Y^T (I - HH^+)Y$ is distributed

$$\chi^2 (n - q, \lambda = \tfrac{1}{2} X^T H^T (I - HH^+) HX) = \chi^2 (n - q, \lambda = 0).$$

This section indicates that in the theory of estimation the generalized inverse seems particularly applicable. It appears that considerable simplification if not amplification of the theory can be made using this tool. Also, it should be noted that separate analysis is not needed to study the full-rank or less-than full-rank regression model. In the case of the full-rank model all functions of the state vector are estimable while the class of estimable functions is restricted in the less than full-rank case. It is also of interest to observe that it is always possible to reparameterize the linear model $Y = HX + V$ such that the new regression matrix is of full-rank. To be specific, suppose the linear model is less than full-rank. Determine a maximal set of linearly independent columns of the regression matrix H. Suppose the maximal set contains k column vectors. Then use these k column vectors to be the first k columns in the new regression matrix. Also, the elements of the state vector X should be rearranged according to the rearrangement in the regression matrix H. Hence one can now write $Y = H_1H_2\tilde{X} + V$ where H_1 consist of the k linearly independent columns, H_2 the remaining columns of H, and $\tilde{X}$ the rearranged elements of X. If one lets $T = H_2\tilde{X}$, then T becomes an estimable function.

5.8. Sequential Least Squares Parameter Estimation

In this section a sequential algorithm for least squares estimation of a parameter state vector is developed utilizing the properties of the Penrose pseudo inverse. This algorithm allows the estimation to begin after the first observation has been made and requires no apriori knowledge of the initial state of the system. The problems of weighted least squares, deleting a bad observation, and application to a dynamical system are also considered. The problems associated with singular matrices encountered in iterative least squares procedures do not affect the algorithm.

Nonlinear parameter estimation problems are usually handled by linear approximations of the actual parameter state in a neighborhood of a nominal parameter state. The resulting equations are of the same general form $Ax = b$; however, in this case x denotes the deviation from the nominal state, and b denotes the deviation in "observed" and "computed" values.

The problem then is to find the solution for x in the matrix equation $Ax = b$, where A is an n by m matrix, x is an m by 1 parameter state vector, and b is an n by 1 observation vector. Since this equation, in general, does not have a solution the normal form is considered $A^{+}Ax = A^{*}b$. It may be shown that this equation always has a solution, in fact has infinitely many solutions when the matrix $A^{*}A$ is singular.

A sequential method for computing the least squares estimate after $n + 1$ observations have been made without having to begin again from the beginning is especially desirable in real time operation. This method allows one to move from the n^{th} to the $(n + 1)^{st}$ parameter state with a minimum of computations.

The problem of the matrix A^TA becoming singular does not affect this algorithm because the solution for x which has minimum Euclidean norm is chosen and the estimation procedure continues on. At the time the matrix A^TA becomes nonsingular this method gives the same solution as the conventional method. One of the best applications of this method is to the problem of orbit determination. In this case one operates in the mode of the deviation space but basically the problem is the same.

There are in existence two well known procedures for attacking this problem. One method, by P. A. Gainer (43) requires that enough observations be made for the system to be fully determined before the procedure begins, another, by R. E. Kalman (56, 57) which requires apriori knowledge of the covariance of the estimate. The method which will be outlined does not require the system to be fully determined before the estimation begins nor does it require any apriori knowledge of the initial state of the system. It will be easily seen that this method becomes the same as Kalman's after the system becomes fully determined.

This section is divided into five major divisions. The first division exhibits the sequential algorithm. The second examines the problem of weighting. The third exhibits the covariance matrix. The fourth presents a method for deleting a bad observation. The fifth demonstrates the application of the sequential algorithm to a dynamical system.

<u>The Sequential Algorithm</u>

Utilizing the properties of the Penrose pseudo inverse, an algorithm is developed which allows one to move sequentially from the n^{th} to the $(n + 1)^{st}$ parameter state with a minimum of operations. This algorithm is not restricted to scalar observations but also allows vector valued observations.

<u>Theorem 5.19</u>: The pseudo inverse of any matrix,

$$A = (U, V) \tag{1}$$

where U and V are arbitrary partitions of the matrix A in columns, can be written in the following form:

$$A^+ = \begin{bmatrix} U^+ - U^+VC^+ - U^+V(I - C^+C)\,KV^TU^{+T}U^+(I - VC^+) \\ \\ C^+ + (I - C^+C)\,KV^TU^{+T}U^+(I - VC^+) \end{bmatrix} \tag{2}$$

where $\quad C = (I - UU^+)V$

and

$$K = (I - (I - C^+C)\,V^TU^{+T}U^+V(I - C^+C)^{-1}$$

The proof to this theorem is given in chapter 3 and is stated here for easy reference.

<u>Corollary 5.1</u>: The pseudo inverse of any matrix,

$$A = \begin{bmatrix} R \\ S \end{bmatrix}$$

where R and S are arbitrary partitions of the matrix A. In rows, can be written in the following form:

$$A^{+} = (R^{+} - JSR^{+} \;\vdots\; J)$$

where

$$J = E^{+} + (I - E^{+}S)R^{+}R^{+T}S^{T}K\,(I - EE^{+})\ ,$$

$$K = (I + (I - EE^{+})SR^{+}R^{+T}S^{T}\,(I - EE^{+})\,)^{-1}\ ,$$

and

$$E = S\,(I - R^{+}R)$$

<u>Proof</u>: Taking the transpose of equations (1) and (2):

$$A^{+T} = \begin{bmatrix} U^{T} \\ V^{T} \end{bmatrix} \tag{3}$$

$$\begin{aligned} A^{+T} = {} & (U^{+T} - C^{+T}V^{T}U^{+T} - (I - C^{+T}V^{T})U^{+T}U^{+}VK^{T}\,(I - C^{T}C^{+T})V^{T}U^{+T} \\ & \vdots\; C^{+T} + (I - C^{+T}V^{T})U^{+T}U^{+}VK^{T}\,(I - C^{T}C^{+T})\,) \end{aligned} \tag{4}$$

with

$$C^{T} = V^{T}\,(I - U^{+T}U^{T})$$

and

$$K^{T} = (I + (I - C^{T}C^{+T})V^{T}U^{+T}U^{+}V\,(I - C^{T}C^{+T})\,)^{-1}\ .$$

Noting that equation (3) is of the desired form, let

$$A = A^{T},\ R = U^{T},\ S = V^{T},\ K = K^{T},\ E = C^{T}$$

and using the fact that $(A^{+})^{+} = A$, gives the desired result.

$$A^+ = [R^+ - E^+SR^+ - (I - E^+S)R^+R^{+T}S^TK\ (I - EE^+)\ SR^+$$

$$\vdots \quad E^+ + (I - E^+S)R^+R^{+T}S^TK\ (I - EE^+)]$$

with

$$E = S\ (I - R^+R)$$

and

$$K = (I + (I - EE^+)SR^+R^{+T}S^T\ (I - EE^+)\ )^{-1}$$

Simplifying, let

$$J = E^+ + (I - E^+S)R^+R^{+T}S^TK\ (I - EE^+)$$

which gives,

$$A^+ = (R^+ - JSR^+ \vdots\ J)$$

It should be noted that K, the inverse of a positive definite matrix, exists for every R and S.

In least squares parameter estimation one encounters the problem of finding the minimum norm solution for x of the matrix equation.

$$Ax = b + e$$

where A is a $n \times m$ matrix, c is a $m \times 1$ parameter state vector, b is an observation vector, and e is an error vector. The least squares solution of minimal Euclidean norm is given by

$$\hat{x} = A^+b$$

<u>Theorem 5.20</u>: Let A be any $n \times k$ matrix, A_{n-1} be the matrix consisting of the first $n-1$ submatrices, and A_n the n^{th} submatrix, that is

$$A = \begin{bmatrix} A_{n-1} \\ \\ A_n \end{bmatrix}$$

<u>Proof</u>: It is noted that A is in the form specified in Corollary 5.1. This implies that:

$$A^+ = (A_{n-1} + J_n A_n A^+_{n-1} \vdots J_n)$$

where

$$J_n = E_n^+ + (I - E_n^+ A_n) A^+_{n-1} A^{+T}_{n-1} A_n^T K_n (I - E_n E_n^+)$$

$$K_n = (I - E_n E_n^+) A_n A^+_{n-1} A^{+T}_{n-1} A_n^T (I - E_n E_n^+))^{-1}$$

and

$$E_n = A_n (I - A^+_{n-1} A_{n-1})$$

Using the results of the above theorem the least squares solution of

$$Ax = b + e$$

may be realized as a sequential process. Noting that the least squares solution after n observations have been made is given by

$$\hat{x}_n = A^+ b$$

and writing A and b in partitioned form

$$x_n = (A^+_{n-1} - J_n A_n A^+_{n-1} \vdots J_n) \begin{bmatrix} b_{n-1} \\ \cdots \\ b_n \end{bmatrix}$$

where b_{n-1} is the first $n - 1$ observation vectors and b_n is the n^{th} observation vector. Multiplying gives

$$x_n = A^+_{n-1} b_{n-1} - J_n A_n A^+_{n-1} b_{n-1} + J_n b_n$$

and noting that $\hat{x}_{n-1} = A^+_{n-1} b_{n-1}$ yields

$$\hat{x}_n = \hat{x}_{n-1} + J_n (b_n - A_n \hat{x}_{n-1})$$

It should be noted that no apriori knowledge is necessary to begin the estimation procedure. One need note only that

$$\hat{x} = A^+_1 b_1$$

to start the procedure. In order to carry out this procedure it is only necessary to compute sequentially the two matrices $A^+_n A_n$ and $A^+_n A^{+T}_n$. Of these two matrices $A^+_n A^{+T}_n$ will later be shown to be the covariance matrix. To compute $A^+_n A_n$,

$$A^+_n A_n = (A^+_{n-1} - J_n A_n A^+_{n-1} \vdots J_n) \begin{bmatrix} A_{n-1} \\ \cdots \\ A_n \end{bmatrix}$$

then

$$A^+_n A_n = A^+_{n-1} A_{n-1} + J_n A_n (I - A^+_{n-1} A_{n-1})$$

and to compute $A_n^+ A_n^{+T}$

$$A_n^+A_n^{+T} = A_{n-1}^+A_{n-1}^{+T} - J_nA_nA_{n-1}^+ \vdots J_n) \qquad \begin{bmatrix} A_{n-1}^{+T} - A_{n-1}^{+T} A_n^T J_n^T \\ \cdots \\ J_n^T \end{bmatrix}$$

or

$$A_n^+A_n^{+T} = A_{n-1}^+A_{n-1}^{+T} - A_{n-1}^+A_{n-1}^{+T}A_n^T J_n^T - J_nA_nA_{n-1}^+A_{n-1}^{+T}$$
$$+ J_nA_nA_{n-1}^+A_{n-1}^{+T}A_n^TJ_n^T + J_nJ_n^+$$

Weighted Observations

By weighted least squares it is meant to minimize $(Ax - b)^TR^{-1}$ $(Ax - b)$, where R is a diagonal positive definite matrix and hence there exists a matrix Q such that $Q^TQ = R^{-1}$. In order to do this it is only necessary to consider a matrix equation of the form $QAx = Qb \pm Qe$ instead of the equation $Ax = b + e$. This indicates that the least squares solution for x is given by,

$$\hat{x} = (QA)^+ Qb \tag{6}$$

Theorem 5.21: A method for computing sequentially the least squares estimate for x in the matrix equation $QAx = Qb + Qe$ when the observations are scalars is given by:

$$\hat{x}_n = \hat{x}_{n-1} + P_n (b_n - a_n\hat{x}_{n-1})$$

where

$$P_n = \begin{cases} (a_n (I - A^+_{n-1} A_{n-1}))^+ & \text{if} \quad a_n \neq a_n A^+_{n-1} A_{n-1} \\ (r_n + a_n A^+_{n-1} A^{+T}_{n-1} a^T_n)^{-1} A^+_{n-1} A^{+T}_{n-1} a^T_n & \text{if} \quad a_n = a_n A^+_{n-1} A_{n-1} \end{cases}$$

<u>Proof</u>: Consider the matrix equation partitioned in the following way:

$$\begin{bmatrix} Q_{n-1} \\ \cdots \\ q_n \end{bmatrix} \begin{bmatrix} A_{n-1} \\ \cdots \\ a_n \end{bmatrix} x = \begin{bmatrix} Q_{n-1} \\ \cdots \\ q_n \end{bmatrix} \begin{bmatrix} b_{n-1} \\ \cdots \\ b_n \end{bmatrix} + \begin{bmatrix} Q_{n-1} \\ \cdots \\ q_n \end{bmatrix} \begin{bmatrix} e_{n-1} \\ \cdots \\ e_n \end{bmatrix}$$

where Q_{n-1}, A_{n-1}, and b_{n-1} are the first n-1 rows of the respective matrices with q_n, a_n, and b_n the n^{th} rows. A simple consequence of Theorem 5.20 indicates that when the observations are scalar valued the sequential least squares solution is given by

$$\hat{x}_n = \hat{x}_{n-1} + s_n q_n (b_n - a_n \hat{x}_{n-1})$$

with

$$(q_n a_n (I - (Q_{n-1}A_{n-1})^+ (Q_{n-1}A_{n-1})))^+$$

if

$$q_n a_n \neq q_n a_n (Q_{n-1} A_{n-1})^+ Q_{n-1}A_{n-1}$$

$$S_n = (1 + q_n a_n (Q_{n-1}A_{n-1})^+ (Q_{n-1}A_{n-1})^{+T} a_n^T q_n^T)^{-1} (Q_{n-1}A_{n-1})^+$$
$$(Q_{n-1}A_{n-1})^{+T} a_n^T q_n^T$$

if

$$q_n a_n = q_b a_b (Q_{n-1}A_{n-1})^+ Q_{n-1}A_{n-1}$$

To begin the estimation procedure

$$\hat{x}_1 = (q_1 a_1)^+ q_1 b_1$$

but since q_1 is a row of the Q matrix composed of only one element which is not zero this is equivalent to multiplying by a scalar. Thus $(q_1 a_1)^+ = a_1^+ q_1^{-1}$ implies that

$$\hat{x}_1 = a_1^+ b_1 \tag{7}$$

Since Q_{n-1} is nonsingular $(Q_{n-1}A_{n-1})^+ Q_{n-1}A_{n-1} = A_{n-1}^+ A_{n-1}$ which reduces the consideration to examining whether

$$a_n = a_n A_{n-1}^+ A_{n-1} \qquad \text{or} \qquad a_n \neq a_n A_{n-1}^+ A_{n-1}$$

If $a_n \neq a_n A_{n-1}^+ A_{n-1}$

then

$$s_n = (q_n a_n (I - A_{n-1}^+ A_{n01}))^+ ,$$ but by the same argument

as used to obtain equation (7)

$$s_n = (a_n (I - A_{n-1}^+ A_{n-1}) a_n^{-1}$$

which implies that

$$\hat{x}_n = \hat{\ }_{n-1} + (a_n (I - A_{n-1}A_{n-1}))^{+} q_n^{-1} q_n (b_n - a_n\hat{x}_{n-1})$$

and let

$$P_n = (a_n (I - A_{n-1}^{+} A_{n-1}))^{+}$$

then

$$\hat{x}_n = \hat{x}_{n-1} + P_n (b_n - a_n \hat{x}_{n-1})$$

If

$$a_n = a_n A_{n-1}^{+} A_{n-1} \quad \text{and defining}$$

$$C_{n-1} = (Q_{n-1} A_{n-1})^{+} (Q_{n-1} A_{n-1})^{+T}$$

then

$$s_n = (1 + q_n a_n C_{n-1} a_n^T q_n^T)^{-1} C_{n-1} a_n^T q_n^T$$

and letting

$$r_n^{-1} = q_n^T q_n \quad ,$$

then

$$s_n = (r_n + a_n C_{n-1} a_n^T)^{-1} C_{n-1} a_n q_n^{-1}$$

which gives $\hat{x}_n = \hat{x}_{n-1} + (r_n + a_n C_{n-1} a_n^T) C_{n-1} a_1^T q_n^{-1} q_n (b_n - a_n \hat{x}_{n-1})$

and let

$$P_n = (r_n + a_n C_{n-1} a_n^T)^{-1} C_{n-1} a_n^T$$

then

$$\hat{x}_n = \hat{x}_{n-1} + p_n (b_n - a_n \hat{x}_{n-1})$$

The calculation of C_n is done in the following way:

$$C_n = (A^+_{n-1} - s_n q_n a_n A^+_{n-1} \vdots s_n) \begin{bmatrix} A^{+T}_{n-1} - A^{+T}_{n-1} a^T Q^T s^T \\ \cdots \\ s^T_n \end{bmatrix}$$

$$C_n = C_{n-1} - s_n q_n a_n C_{n-1} + s_n (a_n C_{n-1} a^T_n + r_n) s^T_n - C_{n-1} a^T_n q^T_n s^T_n$$

If

$$a_n \neq a_n A^+_{n-1} A_{n-1}$$

$$C_n = C_{n-1} - p_n a_n C_{n-1} + p_n (a_n C_{n-1} a^T_n + r_n) p^T_n - C_{n-1} a^T_n p^T_n$$

and if

$$a_n = a_n A^+_{n-1} A_{n-1}$$

$$C_n = C_{n-1} - p_n a_n C_{n-1}$$

<u>Theorem 5.22</u>: A general expression for sequential estimation of a parameter state vector using weighted vector valued observations is:

$$x_n = x_{n-1} - E^+_n Q_n A_n x_{n-1} - (I - E^+_n Q_n A_n) C_{n-1} A^T_n Q^T_n K_n (I - E_n E^+_n) Q_n A_n x_{n-1}$$

$$+ E^+_n Q_n b_n + (I - E^+_n Q_n A_n) C_{n-1} A^T_n Q^T_n K_n (I - E_n E^+_n) Q_n b_n$$

where

$$E^+_n = (Q_n A_n (I - A^+_{n-1} A_{n-1})^+$$

$$K_n = (I + (I - E_nE_n^+)Q_nA_nC_{n-1}A_n^+Q_n^T (I - E_nE_n^+))^{-1}$$

and

$$C_{n-1} = (Q_{n-1}A_{n-1})^+ (Q_{n-1}A_{n-1})^{+T}$$

The proof is a direct application of Theorem 5.20 and expansion as in Theorem 5.21.

A more interesting consequence of this theorem is when it is assumed that $A_n = A_nA_{n-1}^+A_{n-1}$ or $A_n \neq A_nA_{n-1}^+A_{n-1}$. By this it is meant that each of the rows of the matrix A is linearly independent until the system becomes fully determined. In this case, $A_n \nsubseteq A_nA_{n-1}^+A_{n-1}$ and after the system becomes fully determined $A_n = A_nA_{n-1}^+A_{n-1}$. Under this assumption $E_nE_n^+ = I$ or $E_nE_n^+ = \phi$. The fact that $E_nE_n^+ = I$ if $A_n \neq A_nA_{n-1}^+A_{n-1}$ is shown to be true by noticing that $E_nE_n^+$ is an orthogonal projection operator on $R(E_n)$ and since $A_n \neq A_nA_{n-1}^+A_{n-1}$ it is implied that the range space of E_n is all Euclidean space and the identity on Euclidean space is I. If $A_n = A_nA_{n-1}^+A_{n-1}$ then $E_n = E_n^+ = \phi$, which implies that $E_nE_n^+ = \phi$. If $A_n \neq A_nA_{n-1}^+A_{n-1}$, then

$$\hat{x}_n = \hat{x}_{n-1} + E_n^+Q_n (b_n - A_n\hat{x}_{n-1})$$

but
$$E_n^+Q_n = (Q_nA_n (I - A_{n-1}^+A_{n-1}))^+ Q_n$$

and since the rows of $A_n (I - A_{n-1}^+A_{n-1})$ are linearly independent then $Q_nA_n (I - A_{n-1}^+A_{n-1})$ is equivalent to multiplying each row by a set of constants and thus

$$E_n Q_n = (A_n (I - A^+_{n-1} A_{n-1}))^+$$

If you let

$$P_n = (A_n (I - A^+_{n-1} A_{n-1}))^+$$

then

$$\hat{x}_n = \hat{x}_{n-1} + P_n (b_n - A_n \hat{x}_{n-1})$$

If $\quad A_n = A_n A^+_{n-1} A_{n-1}$

then

$$\hat{x}_n = \hat{x}_{n-1} - C_{n-1} A^T_n Q^T_n K_n Q_n A_n \hat{x}_{n-1} + C_{n-1} A^T_n Q^T_n K_n Q_n b_n$$

with

$$K_n = (I + Q_n A_n C_{n-1} A^T_n Q^T_n)^{-1}$$

and

$$C_{n-1} = (Q_{n-1} A_{n-1})^+ (Q_{n-1} A_{n-1})^{+T}$$

Noting that

$$K_n = Q^{T-1}_n (R_n + A_n C_{n-1} A^T_n)^{-1} Q^{-1}$$

which leads to

$$\hat{x}_n = \hat{x}_{n-1} + C_{n-1} A^T_n Q^{T-1}_n (R_n + A_n C_{n-1} A^T_n)^{-1} Q^{-1}_n Q_n (b_n - A_n \hat{x}_{n-1})$$

and letting

$$P_n = C_{n-1} A^T_n (R_n + A_n C_{n-1} A^T_n)^{-1}$$

then

$$\hat{x}_n = \hat{x}_{n-1} + P_n (b_n - A_n \hat{x}_{n-1})$$

Calculating C_{n-1} is done as before. If $A_n \neq A_n A_{n-1}^{+} A_{n-1}$

$$C_n = C_{n-1} - P_n A_n C_{n-1} + P_n (A_n C_{n-1} A_n^T + R_n) P_n^T - C_{n-1} A_n^T P_n^T$$

and if $A_n = A_n A_{n-1}^{+} A_{n-1}$ then

$$C_n = C_{n-1} - P_n A_n C_{n-1}$$

Summarizing the results of this derivation:

$$\hat{x}_n = \hat{x}_{n-1} + P_n (b_n - A_n \hat{x}_{n-1})$$

$$P_n = \begin{cases} (A_n (I - A_{n-1}^{+} A_{n-1}))^{+} & \text{if } A_n \neq A_n A_{n-1}^{+} A_{n-1} \\ \\ C_{n-1} A_n^T (R_n + A_n C_{n-1} A_n^T)^{-1} & \text{if } A_n = A_n A_{n-1}^{+} A_{n-1} \end{cases}$$

$$C_n = \begin{cases} C_{n-1} - P_n A_n C_{n-1} + P_n (A_n C_{n-1} A_n^T + R_n) P_n^T & \\ \qquad - C_{n-1} A_n^T P_n^T & \text{if } A_n \neq A_n A_{n-1}^{+} A_{n-1} \\ \\ C_{n-1} - P_n A_n^T P_n^T & \text{if } A_n = A_n A_{n-1}^{+} A_{n-1} \end{cases}$$

<u>The Covariance of the Estimate</u>

Consider the vector equation $Qb = QAx + Qe$ with A, b, Q, x, and e defined as before where it is assumed that $E(Qe) = \phi$ and that $E(Qee^TQ^T) = I$ where E is the expected value operator. Define $C(\hat{x}_n, \hat{x}_n)$ to be the covariance matrix of the estimate, then

$$C(\hat{x}_n, \hat{x}_n) = E(\hat{x}_n - x)(\hat{x}_n - x)^T).$$

As before the minimum norm solution to the matrix equation $Qb = QAx + Qe$ which is $x = (QA)^+(Qb - Qr)$, is given by $\hat{x}_n = (QA)^+Qb$, which leads to:

$$\begin{aligned}
C(\hat{x}_n, \hat{x}_n) &= E(((QA)^+ Qb - x)((QA)^+ Qb - x)^T) \\
&= E(((QA)^+ Qe)((QA)^+ Qe)^T) \\
&= E((QA)^+ Qee^TQ^T (QA)^{+T}) \\
&= (QA)^+E(Qee^TQ^T)(QA)^{+T} \\
&= (QA)^+(QA)^{+T} = ((QA)^TQA)^+ \\
C(\hat{x}_n, \hat{x}_n) &= (A^TR^{-1}A)^+
\end{aligned}$$

Deleting an Observation

Suppose the n^{th} observation has been made and it is then determined that the observation is bad. It would be desirable to be able to back up and delete this bad observation and then continue on in the estimation procedure without having to begin again from the beginning.

Theorem 5.23: Consider an arbitrary matrix $A = (U, V)$, and assume A^+ is known. Partition A^+ as $A^+ = \begin{matrix} G \\ H \end{matrix}$ where G and H have the same dimension as U^T and V^T, respectively. Then an expression for U^+ in terms of G and H and related matrices is:

$$U^+ = G(I + V(I - HV)^+H)(I - M^+M) \tag{9}$$

with $\quad M = H - (I - HVXI - HV)^+H$

The proof of this theorem is given in chapter 3 and stated again here for easy reference.

Corollary 5.2: Consider an arbitrary matrix $A = \begin{matrix} R \\ S \end{matrix}$ and assume A^+ is known. Partition $A^+ = (F \vdots W)$ where F and W have the same dimension as R^T and W^T, respectively. Then

$$R^+ = (I - BB^+)\ (I + W(I - SW)^+\ S)\ F \tag{10}$$

with $\quad B = W - W(I - SW)^+\ (I - SW)$

Proof: Taking the transpose of equation (9)

$$U^{+T} = (I - M^TM^{+T})(I + H^T\ (I - HV)^{+T}V^T)G^T$$

and

$$U^{T+} = (I - M^TM^{T+})(I + H^T\ (I - V^TH^T)^+V^T)G^T$$

with

$$M^T = H^T - H^T\ (I - V^TH^T)^+\ (I - V^TH^T)$$

By letting $R = U^T$, $S = V^T$, $F = G^T$, $B = M^T$, and $W = H^T$ gives

$$R^+ = (I - BB^+)(I + W(I - SW)^+ S)F$$

and

$$B = W - W(I - SW)^+ (I - SW)$$

A method for deleting a bad observation which is detected after the next estimate has been calculated is obtained by using the preceding corollary. This is done by partitioning the matrix A and the vector b as before, then the n^{th} estimate is given by

$$\hat{x}_n = \begin{bmatrix} A_{n-1} \\ \cdots \\ A_n \end{bmatrix}^+ \begin{bmatrix} b_{n-1} \\ \cdots \\ b_n \end{bmatrix}$$

Partitioning $A^+ = (F \vdots W)$

then

$$\hat{x}_n = (F \vdots W) \begin{bmatrix} b_{n-1} \\ \cdots \\ b_n \end{bmatrix}$$

and multiplying

$$\hat{x}_n = Fb_{n-1} + Wb_n \qquad (11)$$

The idea is to be able to obtain $\hat{x}_{n-1}$ in terms of x_n, A_n, b_n, and C_n which are all available at the n^{th} observation stage.

Noting that $\hat{x}_{n-1} = A^{+}_{n-1}b_{n-1}$ and substituting into equation (10) then

$$\hat{x}_{n-1} = (I - B_nB_n^{+})(I + W(I - A_nW)^{+}A_n)Fb_{n-1}$$

using equation (11)

$$\hat{x}_n = (I - B_nB_n^{+})(I + W(I - A_nW)^{+}A_n(\hat{x}_n - Wb_n) \tag{12}$$

Since $(A^TA)^{+}A^T = A^{+}$ we have $A_n^{+} = C_nA_n^T$ and by partitioning A gives

$$A_n^{+} = C_n(A^T_{n-1} \vdots A_n^T)$$

which implies that $W = C_nA_n^T$. This substituted into equation (12) leads to

$$\hat{x}_{n-1} = (I - B_nB_n)(I - C_nA_n^T(I - A_nC_nA_n^T)^{+}A_n)(\hat{x}_n - C_nA_n^Tb_n)$$

with

$$B_n = C_nA_n^T - C_nA_n^T (I - A_nC_nA_n^T) \quad (I - A_nC_nA_n^T)$$

and C_{n-1} is obtained as $C_{n-1} = A^{+}_{n-1} A_{n-1}$

Application to a Dynamical System

In nonlinear parameter estimation of a dynamical system one usually uses a linear approximation of the actual parameter state in a neighborhood of a nominal parameter state. In this case the problem of the best estimate for x in the matrix equation $Ax = b + e$ is also encountered. The problem is basically the same, but in this case x denotes the deviation in observed and computed, and A denotes a mapping matrix maltiplied by a state transition matrix. The equation is of the form:

$$\begin{bmatrix} Ht_1\ \phi(t_1, t_0) \\ Ht_2\ \phi(t_2, t_0) \\ \cdot \\ \cdot \\ \cdot \\ Ht_n\ \phi(t_n, t_0) \end{bmatrix} \bar{x} = \begin{bmatrix} y_{t1} \\ y_{t2} \\ \cdot \\ \cdot \\ \cdot \\ y_{tn} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

where Ht_n is a mapping matrix at time t_n, (t_n, t_0) is a state transition matrix from time t_0 to time t_n, $\bar{x}$ is the state vector at time t_0, y_{tn} is the $t_n{}^{th}$ observation vector, and e_n is the n^{th} error vector. This is applied to the method derived in equation (10) and the same assumption is made that $H_{t_n}\phi(t_n, t_0)$ is equal to $H_{t_n}(t_n, t_0)D^+_{t-1}D_{t_n-1}$ or not equal, with

$$D_{t_n} = \begin{bmatrix} Ht_1\ \phi(t_1, t_0) \\ Ht_2\ \phi(t_2, t_0) \\ \cdot \\ \cdot \\ \cdot \\ Ht_n\ \phi(t\ \ , t_0) \end{bmatrix}$$

To make this assumption means that it is assumed that all the rows of D_{t_n} are linearly independent until the system becomes fully determined. This assumption becomes natural in most estimation procedures because the system is of the desired form. The resulting equation for estimating sequentially the deviation in the initial conditions is:

$$\hat{\bar{x}}_{t_n} = \hat{\bar{x}}_{t_n - 1} + P_{t_n} (y_{t_n} - H_{t_n} \phi(t_n, t_0) \hat{\bar{x}}_{t_n - 1}) \tag{13}$$

with

$$P_{t_n} = \begin{cases} (H_{t_n} \phi(t_n, t_0)(I - Dt_{n-1} Dt^{+}_{n-1}))^{+} \\ \qquad \text{if } H_{t_n} \phi(t_n, t_0) \neq H_{t_n} \phi(t_n, t_0) Dt^{+}_{n-1} Dt_{n-1} \\ Ct_{n-1} \phi(t_n, t_0)^T H^T_{t_n} (R_{t_n} + H_{t_n} \phi(t_n, t_0) Ct_{n-1} \phi(t_n, t_0) H^T_{t_n})^{-1} \\ \qquad \text{if } H_{t_n} \phi(t_n, t_0) = H_{t_n} \phi(t_n, t_0) Dt^{+}_{n-1} Dt_{n-1} \end{cases}$$

and

$$C_{t_n - 1} = D^{+}_{t_n - 1} D^{+T}_{t_n - 1}$$

$\hat{\bar{x}}_{t_n}$ is the estimate time t_n of the deviation in the initial state vector and $\hat{x}_{t_n}$ is the estimate at time t_n of the deviation in the state vector of position at time t_n. Since it is desirable to be able to calculate $\hat{\bar{x}}_{t_n}$ it is noted that

$$\hat{\bar{x}}_{t_n} = \phi(t_n, t_{n-1}) \hat{\bar{x}}_{t_n - 1} \tag{14}$$

which leads to

$$\hat{\bar{x}}_{t_n} = \Phi(t_n, t_0)\, \hat{\bar{x}}_{t_0} \tag{15}$$

and

$$\hat{x}_{t_n} = \Phi(t_n, t_0)\, \hat{\bar{x}}_{t_n} \tag{16}$$

Substituting equation (16) into equation (13) leads to,

$$\hat{\bar{x}}_{t_n} = \Phi(t_{n-1}, t_0)^{-1}\hat{x}_{t_n-1} + P_{t_n}(y_{t_n} - H_{t_n}\Phi(t_n, t_0)\Phi(t_{n-1}, t_0)^{-1}\hat{x}_{t_n-1})$$

and substituting equation (16) again

$$\hat{x}_{t_n} = \Phi(t_n, t_{n-1})\hat{x}_{t_n-1} + \Phi(t_n, t_0)P_t\, (y_{t_n} - H_{t_n}\Phi(t_n, t_{n-1})\, \hat{x}_{t_n-1} \tag{17}$$

If $\quad H_{t_n}\Phi(t_n, t_0) \neq H_{t_n}\Phi(t_n, t_0)\, D^{+}_{t_n-1}D_{t_n-1} \quad$ then

C_{t_n} is computed sequentially as before,

$$\begin{aligned} C_{t_n} = {} & C_{t_n-1} - P_{t_n}H_{t_n}\Phi(t_n, t_0)\, C_{t_n-1} \\ & + P_{t_n}(H_{t_n}\Phi(t_n, t_0)\, C_{t_n-1}\Phi(t_n, t_0)^T H^T_{t_n} + R_{t_n})P^T_{t_n} \\ & - C_{t_n-1}\Phi(t_n, t_0)^T H^T_{t_n} P^T_{t_n} \end{aligned}$$

and

$$D^{+}_{t_n}D_{t_n} = D^{+}_{t_n-1} + P_{t_n}P^{+}_{t_n}$$

If

$$H_{t_n}\Phi(t_n, t_0) = H_{t_n}\Phi(t_n, t_0)D^{+}_{t_n-1}D_{t_n-1}$$

then

$$\hat{x}_{t_n} = \Phi(t_n, t_{n-1})\hat{x}_{t_n-1} + L_{t_n}(y_{t_n} - H_{t_n}\Phi(t_n, t_{n-1})\hat{x}_{t_n-1}) \tag{18}$$

with

$$L_{t_n} = G_{t_n-1}H^T_{t_n}(R_{t_n} + H_{t_n}G_{t_n}H^T_{t_n})^{-1}$$

$$G_{t_n} = H^+_{t_n}H^{+T}_{t_n}$$

$$G_{t_n} = G_{t_n-1} - L_{t_n}H_{t_n}G_{t_n-1}$$

and

$$D^+_{t_n}D_{t_n} = D^+_{t_n-1}D_{t_n-1}$$

Equation (18) is obtained by noting that

$C_{t_n} = D^+_{t_n}D^{+T}_{t_n} = (D^T_{t_n}D_{t_n})^+$ where t_n is the time when the system becomes fully determined. At this point in time the generalized inverse becomes the normal inverse and the reversal rule may be used.

$$C_{t_n} = \left(\begin{bmatrix} H_{t_1}\Phi(t_1, t_0) \\ H_{t_2}\Phi(t_2, t_0) \\ \vdots \\ H_{t_n}\Phi(t_n, t_0) \end{bmatrix}^T \begin{bmatrix} H_{t_1}\Phi(t_1, t_0) \\ H_{t_2}\Phi(t_2, t_0) \\ \vdots \\ H_{t_n}\Phi(t_n, t_0) \end{bmatrix} \right)^{-1}$$

$$C_{t_n} = (\Phi(t_1, t_0)^T H^T_{t_0} H_{t_1}\Phi(t_1, t_0) + \Phi(t_2, t_0)^T H^T_{t_2} H_{t_2}\Phi(t_2, t_0)$$
$$+ \ldots + \Phi(t_n, t_0)^T H^T_{t_n} H_{t_n}\Phi(t_n, t_0))^{-1}$$

$$\begin{aligned} C_{t_n} = \Phi(t_n, t_0)^{-1}(&\Phi(t_n,t_1)^{T-1}H_{t_1}^T H_{t_1}\Phi(t_n, t_1)^{-1} \\ &+ \Phi(t_n, t_2)^{T-1}H_{t_2}^T H_{t_2}\Phi(t_n, t_2)^{-1} + \ldots + \\ &\Phi(t_n, t_{n-1})^{T-1}H_{t_n-1}^T H_{t_n-1}\Phi(t_n, t_{n-1})^{-1} + H_{t_n}^{+} H_{t_n})\Phi(t_n,t_0)^{T-1} \end{aligned}$$

and substituting into equation (18). Since the noted inverses are to be taken of state transition matrices and this class of matrix is implectic then the inverse is easily calculated without normal matrix inversion procedures.

Conclusions

In real time operations one encounters the problem of estimating sequentially a parameter state vector. The preceding derivation outlines a procedure which one may use in moving from the n^{th} observation state to the $(n + 1)^{st}$ observation state whether one has vector valued or scalar valued observations. This section presents a way of deleting a bad observation and studies the application to a dynamical system. It is believed that this method will produce more accurate results than the methods now being used which were mentioned earlier in this section. It should also be noted that this technique can be modified slightly so that in the sequential procedure any number of observations can be made before the next estimate is calculated.

5.9 A Generalization of the Wielandt Inequality

The spectral condition number $K(A)$ of an arbitrary non-singular matrix A is defined in terms of the spectral norm [53. p. 81] by

$$K(A) = ||A|| \cdot ||A^{-1}||.$$

The condition number serves two closely related purposes;

(1) it is an index of near singularity, hence can provide a measure of the computational instability to be expected in the process of inversion,

(2) by means of the Kantorovich inequality it provides a measure of the rate of convergence of certain iterative processes [53., p. 100].

Clearly, if A is an arbitrary singular matrix $K(A)$ would not be finite; however, one can define quite naturally a generalization of the spectral condition number by defining for square matrices

$$K(A) = ||A^{+}|| \quad ||A||$$

The purpose of this section is to develop generalizations of the Wielandt and Kantorovich inequalities using a generalized spectral condition number. A discussion of these inequalities for non-singular matrices can be found in [53., pp. 81-84].

Let $||x||$ denote the ordinary Euclidean norm of the vector x, and let the associated operator norm be

$$||A|| = \sup_{||x|| = 1} ||Ax||.$$

A singular value of a matrix A is the non-negative square root of an eigenvalue of AA^*, where * indicates the conjugate transpose of A. Thus the value of $||A||$ is the largest singular value of A. We will denote the singular values of A by $\sigma_i(A)$ and the eigenvalues of A by $\lambda_i(A)$.

We now define formally what is meant by a generalized spectral number of an arbitrary square matrix.

<u>Definition 5.4</u>: <u>The generalized spectral condition number</u> $K(A)$ <u>of an arbitrary square matrix</u> A <u>is defined by the equality</u>

$$K(A) = ||A|| \cdot ||A^+||$$

The following properties can readily be established:

(P1) $K(A) \geq 1$,

(P2) $K(A) = K(A^+) = K(A^*)$,

(P3) if $||(AB)^+|| \leq ||A^+||\ ||B^+||$ then $K(AB) \leq K(A)K(B)$.

The symbols $R(A)$ and $N(A)$ will denote the range and null space of A, respectively and a superscript $\perp$ will denote their orthogonal complements.

It is convenient to state and prove the following preliminary lemmas:

<u>Lemma 5.5</u>: <u>Let</u> $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_k^2$ <u>be the eigenvalues of</u> $M = A^*A$ <u>corresponding to eigenvectors in</u> $N(A)^\perp$, <u>then for any unit vector</u> x <u>in</u> $N(A)^\perp$, $\sigma_k^2 \leq x^*Mx$.

Proof: Let $x_1, x_2, \ldots, x_k$ be an orthonormal set of basis vectors for $N(A)^{\perp}$ which are eigenvectors of M corresponding to $\sigma_1^2, \ldots, \sigma_k^2$, respectively. Let $x = \sum_{i=1}^{k} a_i x_i$ so that

$$x^* Mx = \sum_{i,j}^{k} a_i \bar{a}_j \sigma_i^2 x_i^* x_j = \sum_{i}^{k} a_i \bar{a}_i \sigma_i^2$$

$$\geq \sum_{i}^{k} |a_i|^2 \sigma_k^2 = \sigma_k^2$$

Lemma 5.6: Let $G = (x,y)^* M(x,y)$ where x and y are orthonormal vectors in $N(A)^{\perp}$. Then the field of values $F(G)$ is contained in

$$\{w^* Mw \mid w^* w = 1, \ w \varepsilon N(A)^{\perp}, \ M = A^* A\}.$$

Proof: By definition, $F(G)$ is the set of all (complex) scalars of the form $z^* Gz$ as z varies over all possible unit vectors. Let $z^* = (\bar{z}_1, \bar{z}_2)$, and $w = (z_1 x + z_2 y)$, then $z^* Gz = w^* Mw$. Certainly w is a vector in $N(A)^{\perp}$ and $w^* w = 1$, hence, the conclusion follows.

The main result of this section is summarized in the following theorem:

Theorem 5.24: For any square matrix A and any pair of orthonormal vectors x and y in $N(A)$.

$$|x^* My \leq ||Ax|| \quad ||Ay|| \quad \cos\theta,$$

where

$$\cot(\theta/2) = K(A)$$

Proof: The inequality is trivial if A is of rank 1, hence we assume that if A is singular, it is of rank at least two. Now $K(M) = ||A^*A||\ ||(A^*A)^+|| = ||A||^2\ ||A^+||^2 = K^2(A)$, since M is symmetric, and $(M^+)^+ = (M^2)^+$.

Let $x, y \in N(A)^{\perp}$ and form the two dimensional section

$$G = (x, y)^* M(x, y)$$

The rows of G can be shown to be linearly independent, thus G is nonsingular. Let λ_1 and $\lambda_2 \leq \lambda_1$ be the eigenvalues of G and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_K$ the non-zero singular value of A. Then $\sigma_1^2 = \lambda_i(M)$ and $\sigma_1^2 \geq \lambda_1 \geq \lambda_2 \geq \sigma_K^2$. The letter follows from the fact that G is non-singular and lemmas 5.5 and 5.6.

Let $\delta(G)$ denote the determinant of G and consider

$$1 - \frac{|x^*My|^2}{x^*Mxy^*My} = \frac{4\delta(G)}{(x^*Mx+y^*My)^2 - (x^*Mx-y^*My)^2}$$

$$= \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2 - (x^*Mx-y^*My)^2}$$

If x and y are allowed to vary throughout $N(A)^{\perp}$, the right member is minimized and $|x^*My|^2/(x^*Mxy^*My)$ is maximized when $x^*Mx = y^*My$. When this is true, $|x^*My|^2/(x^*Mxy^*My) = (\lambda_1/\lambda_2-1)^2/(\lambda_1/\lambda_2+1)^2$. The right member is a monotonically increasing function of λ_1/λ_2, hence is greatest when $\lambda_1 = \sigma_1^2$ and $\lambda_2 = \sigma_K^2$.

Thus to maximize $|x^*My|^2/(x^*Mxy^*My)$ with respect to all orthonormal pairs of vectors x and y in $N(A)^{\perp}$, these vectors must be taken in the planes of the eigenvectors υ_1 and υ_K of M belonging to the eigenvalues $\sigma_1{}^2$ and $\sigma_K{}^2$, respectively.

Evidently $K(A) = \sigma_1/\sigma_K$. Hence, if the angle θ is defined by

$$\frac{K(A) - K(A)^{-1}}{K(A) + K(A)^{-1}} = \cos \theta$$

we obtain

$$|x^*Mx|/(x^*Mxy^*My) \leq ||Ax|| \; ||Ay|| \cos \theta, \tag{1}$$

the desired generalized Wielandt inequality.

In the case in which A is non-singular, the Kantorovich inequality is simply a special case of the Wielandt inequality. This holds for the generalized situation also.

<u>Corollary 5.3</u>: <u>For any square matrix</u> A <u>and unit vector</u> x <u>in</u> $N(A)^{\perp}$,

$$(x^*x)^2 \geq x^*Mx \; x^*M^+ x \sin^2 \theta.$$

<u>Proof</u>: Let $y = (x^*x)M^+x - (x^*M^+x)x$. It is easily verified that $x^*y = 0$ and that $y \in N(A)^{\perp}$.

Let $u_t = x^* M^t x$ and $u_{-t} = x^*(M^+)^t x$ for $t = 0, 1, 2, \ldots$. Then $y = u_0 M^+ x - u_{-1} x$ and $x^* My = u_0 x^* MM^+ x - u_{-1} u_1$. But $N(A)^\perp = N(M)^\perp = R(M^+) = R(M)$ where $R(M)$ is the range space of M so that $M^+ Mx = x$ and also $MM^+ x = x$ for $x \varepsilon N(A)^\perp$. Hence $x^* My = u_0 - u_{-1} u_1$ and $y^* My = u_1(u_1 u_{-1} - u_0^2)$. Therefore $\delta(G)/(x^* Mxy^* My) = u_0^2 - u_1 u_{-1}$. But (1) implies that this is not less than $\sin^2\theta$ and the corollary, a generalization of the Kantorovich inequality, follows, that is,

$$(x^* x)^2 \geq x^* Mx \; x^* M^+ x \sin^2\theta.$$

An Application

As an indication of the usefulness of the generalizations presented here consider the following concerning the problem of solving a system (that is, obtaining a solution vector x)

$$Ax = h$$

in which we select our initial guess, say x_0 and iterate to an approximate solution. We require that $h \varepsilon R(A)$ which implies that a solution exists.

Let

$$x_{n+1} = x_n + C_n r_n \tag{2}$$

where

$$r_n = h - Ax_n = A(x - x_n) = As_n \tag{3}$$

and

$$s_n = x - x_n.$$

A method of projection is briefly one that assigns at any step a subspace, defined by $P_n \geq 1$ linearly independent columns of matrix Y_n and selects u_n in such a way that if

$$C_n r_n = Y_n u_n,$$

then

$$s_{n+1} = s_n - Y_n u_n$$

is reduced in some norm.

For a given Y_n, it is required to minimize

$$s_{n+1}^* \; s_{n+1} = (s_n - Y_n u_n)^* \; (s_n - Y_n u_n).$$

Let

$$u_n = (Y_n^* Y_n)^+ Y_n^* s_n + w_n \tag{4}$$

where w_n is an arbitrary vector.

Then

$$s_{n+1}^* s_{n+1} = s_n^* s_n - s_n^* Y_n (Y_n^* Y_n)^+ Y_n^* s_n + w_n^* Y_n^* Y_n w_n. \tag{5}$$

Since $Y_n^* Y_n$ is positive definite, (5) is minimized when $w_n = 0$. Therefore,

$$||s_n||^2 - ||s_{n+1}||^2 = s_n^* Y_n (Y_n^* Y_n)^+ Y_n^* s_n. \tag{6}$$

The matrices Y_n, which are, ordinarily, single column vectors must change from step to step. The choice Y_n is equivalent to the choice,

$$C_n A = Y_n (Y_n^* Y_n)^+ Y_n^* \tag{7}$$

since

$$C_n r_n = C_n A s_n = Y_n u_n = Y_n (Y_n^* Y_n)^+ Y_n^* s_n$$

for any s_n. This in turn implies that

$$C_n A = Y_n (Y_n^* Y_n)^+ Y_n^* .$$

But this is feasible only if $Y_n^* = V_n^* A$ for some V_n, since otherwise C_n could be obtained only after calculating A^+.

Taking $Y_n^* = V_n^* A$, where V_n^* is selected and Y_n computed we have

$$C_n A = A^* V_n (V_n^* AA^* V_n)^+ V_n^* A$$

or

$$C_n = A^* V_n (V_n^* AA^* V_n)^+ V_n^* AA^+ . \tag{8}$$

It follows then from (6), (7), and (8) that

$$||s_n||^2 - ||s_{n+1}||^2 = s_n^* A^* V_n (V_n^* AA^* V_n)^+ V_n^* As_n$$

$$= r_n^* V_n (V_n^* AA^* V_n)^+ V_n^* r_n .$$

If each $V_n = v_n$ is a single column vector, then from (2) and (8)

$$x_{n+1} = x_n + A^* v_n (v_n^* AA^* v_n)^+ v_n^* r_n$$

since $r_n = As_n \varepsilon R(A)$.

Let $u_n = v_n^* r_n/(v_n^* AA^* r_n)$. Then $x_{n+1} = x_n + u_n A^* v_n$ and

$$||s_n||^2 - ||s_{n+1}||^2 = u_n r_n^* v_n ,$$

or

$$(s_{n+1}^* s_{n+1})/(s_n^* s_n) = 1 - |r_n^* v_n|^2/s_n^* s_n v_n^* AA^* v_n. \tag{9}$$

The method of steepest descent imposes the restriction that $V_n = r_n$. The generalized Kantorovich inequality then provides a bound for the right side of (9). By the corollary and (3) it follows that

$$|r_n^* r_n|^2 \geq r_n^* AA^* r_n r_n^* (AA^*)^+ r_n \sin^2\theta$$

$$\geq r_n^* AA^* r_n r_n A^{+*} A^+ r_n \sin^2\theta$$

$$\geq r_n^* AA^* r_n s_n^* s_n \sin^2\theta.$$

And finally from (9) we conclude that

$$|| s_{n+1}||^2/||s_n||^2 \geq \cos^2 \theta.$$

CHAPTER 6

COMPUTATIONAL PROCEDURES

6.1 Introduction

In recent years many techniques for obtaining the pseudo inverse of a matrix have appeared in the literature. These techniques are of various natures, including explicit expressions for each element of the pseudo inverse to approximate determinations. In this chapter several of these methods are presented along with some comment concerning their merits. Each method is identified by the name of the author of the paper in which the method is presented. A numerical example is included to illustrate the method presented. The same example is used to illustrate each method to facilitate comparison of the computation schemes.

6.2 Householder Method

Let A be an arbitrary matrix of rank r, and let $A = FR^*$ where each of F and R have r linearly independent columns. Then $A^+ = R(R^*R)^{-1}(F^*F)^{-1}F^*$. Although this method gives an explicit form for A^+ it requires that: (1) A be factored as FR^*, which is not easily accomplished on an electronic computer. (2) The inverse of the non-singular matrices R^*R and F^*F need to be obtained, and (3) the product needs to be formed. The factorization of A may be accomplished by several procedures. Householder [53] mentions three methods. The first of these is an iterative scheme. Let

$A = A^{(o)} = (a_1^{(o)}, a_2^{(o)}, \ldots, a_m^{(o)}) \neq 0$ be any $n \times m$ matrix and e_k be the k^{th} column of the identity matrix of size $k \times k$. Let j_1 be the smallest index for which $a_{j_1}^{(o)} \neq 0$ and $e_{i_1}^T a_{j_1}^{(o)} \neq 0$. That is, the element in position i_1 is nonnull. Let I_{ij} denote the elementary permutation matrix whose effect of multiplying a matrix on the left by I_{ij} is to interchange the i^{th} and j^{th} rows; that of multiplying on the right is to interchange columns. Then for some elementary triangular matrix L_1^{-1}, possibly the identity, the matrix

$$A^{(1)} = L_1^{-1} I_{1i_1} A^{(o)}$$

is null in the first j_1^{-1} columns, and null below the first element in the next column.

If every row below the first in $A^{(o)}$ is null, the algorithm is complete. If not, and if $m = 2$, let $A^{(2)} = A^{(1)}$ and the algorithm is again complete. Otherwise, repeat the process by picking out the first column containing a nonnull element before the first. Let this be $a_{j_2}^{(1)}$. Then for some i_2, $e_2^T I_{2i_2} A_{j_2}^{(1)} \neq 0$. Hence for some L_2^{-1}, the matrix

$$A^{(2)} = L_2^{-1} I_{ei_2} A^{(1)}$$

is null in the first $j_1 - 1$ columns, null below the first element in the next $j_2 - j_1$ columns, and null below the second element in the next column.

The process is continued until reaching $A^{(p)}$ where either $p = n$ or else all the rows below the p^{th} are null. Hence

$$A = M_1 A^{(p)}$$

where

$$M_1 = I_{1i_1} L_1 I_{2i_2} L_2 \cdot \cdot \cdot$$

which is nonsingular, and $A^{(p)}$ is a matrix whose first rows are linearly independent with the remaining rows null. If $p < n$, let P be the matrix obtained by dropping the $n - p$ rows from $A^{(p)}$ and let M be the matrix obtained by dropping the last $n - p$ columns of M_1. Then $F = M$ and $P^* = R$ will be the required factorization.

The products R^*R and R^*F are Hermitian and full rank so that their inverses can be obtained by standard procedures for inverting symmetric matrices. Finally the pseudo inverse, A^+, is computed as the product

$$A^+ = R(R^*R)^{-1} (F^*F)^{-1} F^*$$

This method requires the factorization of A which, by the method outlined above, requires considerable searching and matrix manipulation although the process is straightforward. The two required inversions are also troublesome, however, there are relatively good procedures for inverting Hermitian matrices and the two inversions could perhaps be done in parallel.

An example is now given to illustrate the above technique.

<u>Example</u>: Let $A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} = A^{(o)}$

which is a 3 x 4 matrix of rank 2.

Let $I_{11} = I$ and

$$L_{11}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad \text{then}$$

$$A^{(1)} = L_{11}^{-1} I_{11} A^{(o)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix}$$

Let $I_{22} = I$,

$$L_{22}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \quad \text{then}$$

$$A^{(2)} = L_{22}^{-1} I_{22} A^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence $p = r(A) = 2$, so that

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix} \text{ and } M_1 = L_1 L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Thus $M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$.

Letting $F = M$ and $P^* = R$ we obtain

$$R^* R = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{and} \quad F^* F = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

Thus

$$A^+ = R\,(R^* R)^{-1} \quad (F^* F)^{-1}\, F^*$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2/5 & 1/5 \\ 1/5 & 3/5 \end{pmatrix} \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$= \quad 1/15 \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & -5 & -1 \\ 3 & 0 & 3 \end{pmatrix} .$$

6.3 Penrose Methods:

A somewhat similar method is proposed by R. Penrose in [71]. His method is to partition the given matrix after suitably arranging the rows and columns in the following form:

$$\begin{pmatrix} A & B \\ C & CA^{-1}B \end{pmatrix}$$

where A is a non-singular submatrix whose rank is equal to that of the whole matrix. Then the pseudo inverse is given by

$$\begin{pmatrix} A & B \\ C & CA^{-1}B \end{pmatrix}^{*} = \begin{pmatrix} A^{*}PA^{*} & A^{*}PC^{*} \\ B^{*}PA^{*} & B^{*}PC^{*} \end{pmatrix}$$

Where $P = (AA^{*} + BB^{*})^{-1} A (A^{*}A + C^{*}C)^{-1}$. The matrices $AA^{*} + BB^{*}$ and $A^{*}A + C^{*}C$ are positive definite, since A is non-singular. Thus, the pseudo inverse of any matrix can be expressed in terms of ordinary inverses of matrices.

The same two basic problems as in Section 2 appear again. Namely, to partition the given matrix into a form containing a sub-matrix with rank equal to the rank of the given matrix, and to obtain the inverses of two nonsingular matrices. Also, a method must be available for arranging the given matrix into the required form. This could be some method of elimination in which the rank is determined at the same time as the arrangement is done.

Using the same matrix as in Section 2 we illustrate this method.

Example: Let

$$R = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$, $C = (1, 1)$, then

$A^{-1} = A$, $CA^{-1}B = (0, 1)$, $A^{*}A = AA^{*} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,

$$BB^* = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad C^*C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} .$$

Hence $P = (AA^* + BB^*)^{-1} \; A(A^*A + C^*C)^{-1}$

$$= \begin{pmatrix} 2/5 & 1/5 \\ 1/5 & 3/5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix} = 1/15 \begin{pmatrix} 3 & 0 \\ -1 & 5 \end{pmatrix} .$$

Hence

$$R^+ = \begin{pmatrix} A^*PA^* & A^*PC^* \\ B^*PA^* & B^*PC^* \end{pmatrix}$$

$$= \begin{pmatrix} 1/15 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & 1/15 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ 1/15 \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & 1/15 \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{pmatrix}$$

$$= 1/15 \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & -5 & -1 \\ 3 & 0 & 3 \end{pmatrix}$$

By comparing this example with that in Section 2, it can be seen that the two methods are equivalent. The two matrices which are inverted are identical.

Method II (Penrose)

Penrose has also given a recursive method for obtaining the pseudo inverse of a matrix which is extremely concise.

He makes use of the fact that the pseudo inverse of a given matrix A can be obtained from any matrix D satisfying $A^*A = D(A^*A)^2$, since multiplication on the right by $A^+A^{+*}A^+$ gives $A^+ = DA^*$. Let $B = A^*A = (b_{ij})$, and define a sequence of matrices C_j for $j = 1, 2, \ldots$ by

$$C_1 = I$$

$$C_{j+1} = I \cdot \frac{1}{j} \text{ trace } (C_jB) - C_jB$$

Penrose has shown that if r is the rank of the matrix A, then $C_{r+1}B = 0$ and trace $(C_rB) \neq 0$, so that is $D = \dfrac{rC_r}{\text{trace } (C_rB)}$, then $DB^2 = B$ as required.

This method has many advantages: It does not require the factorization of any matrix; it does not necessitate taking the inverse of a matrix. It involves only the basic matrix operations of scalar multiplication, matrix addition, multiplication and the trace. It does <u>not</u> require initial knowledge of the rank of A, but rather this is determined by the algorithm since the iteration stops when trace $(C_kB) = 0$, which is the case when $K = \text{tr } (A) + 1$.

If A is $n \times m$, the procedure can be adapted so that the matrices B, C_i, D are all square of size min (n, m). Hence, if $m \leq n$, $B = A^*A$ is $n \times n$ and the method is used as outlined. However, if $m > n$, the pseudo inverse of A^* is found so that the dimensions of B are $n \times n$. Then using the relation $(A^*)^+ = (A^+)^*$, the pseudo inverse of A is obtained.

Example: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} .$$

Then

$$B = A^*A = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix} .$$

Now $C_1 = I$, $C_1B = B$ and

$$C_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \frac{(2+2+2+2)}{1} - \begin{pmatrix} 2 & 1 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 6 & -1 & -1 & -2 \\ -1 & 6 & 1 & -1 \\ -1 & 1 & 6 & -1 \\ -2 & -1 & -1 & 6 \end{pmatrix}$$

$$C_2B = \begin{pmatrix} 6 & 3 & 3 & 6 \\ 3 & 9 & -6 & 3 \\ 3 & -6 & 9 & 3 \\ 6 & 3 & 3 & 6 \end{pmatrix} .$$

Thus

$$C_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \frac{(6+9+9+6)}{2} - \begin{pmatrix} 6 & 3 & 3 & 6 \\ 3 & 9 & -6 & 3 \\ 3 & -6 & 9 & 3 \\ 6 & 3 & 3 & 6 \end{pmatrix}$$

$$= \begin{pmatrix} 9 & -3 & -3 & -6 \\ -3 & 6 & 6 & -3 \\ -3 & 6 & 6 & -3 \\ -6 & -3 & -3 & 9 \end{pmatrix} .$$

$C_3B = 0$, thus the rank of A is 2, the iteration ceases, and A^+ is given by

$$A^+ = DA^* = \frac{rC_rA^*}{tr(C_rB)} = \frac{2}{30} \begin{pmatrix} 6 & -1 & -1 & -2 \\ -1 & 6 & 1 & -1 \\ -1 & 1 & 6 & -1 \\ -2 & -1 & -1 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

or

$$A^+ = \frac{1}{15} \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & -5 & -1 \\ 3 & 0 & 3 \end{pmatrix}$$

To show that this method is programmable for computer solution, a flow chart is included.

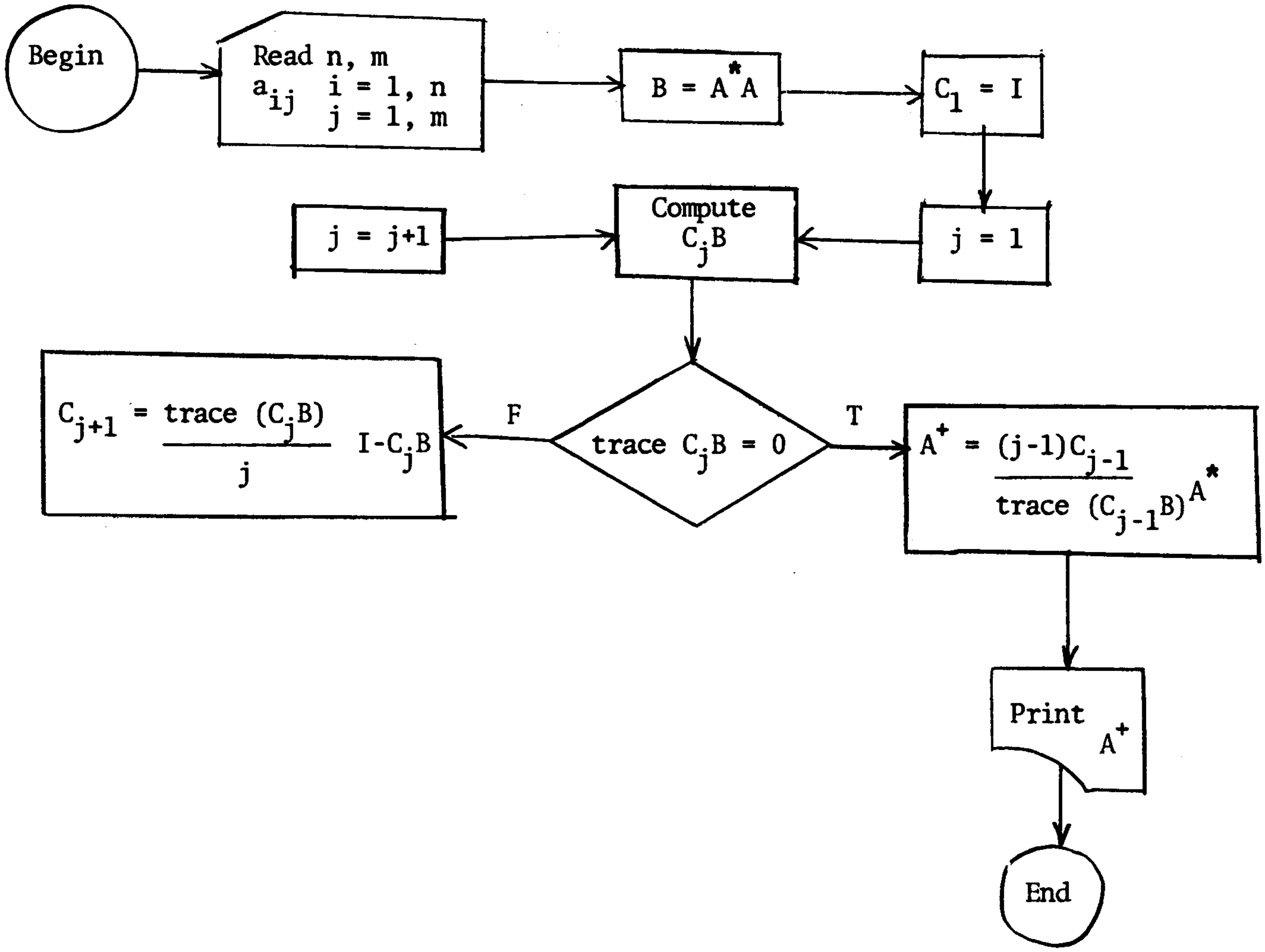

Fig. 1 Flow Chart for Penrose Method 2.

6.4 The Decell (Cayley-Hamilton) Algorithm I

A formula for the pseudo inverse of an arbitrary complex matrix has been derived by Decell [33] based upon the Cayley-Hamilton Theorem. This computing scheme leads directly to a recursive algorithm which is easily adapted to machine computation.

Theorem 6.1: Let A be any $n \times m$ complex matrix and let $f(\lambda) = (-1)^n(a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_k\lambda^{n-k} + \ldots + a_{n-1}\lambda + a_n I)$ be the monic $(a_0 = 1.)$ characteristic polynomial of AA^*. If $1 \neq 0$ is the largest integer such that $a_k \neq 0$ then the pseudo inverse A^+ of A is given by

$$A^+ = -a_k^{-1}A^*[(AA^*)^{k-1} + a_1(AA^*)^{k-2} + \ldots + a_{k-2}(AA^*) + a_{k-1}I] .$$

If $k = 0$ is the largest integer for which $a_k = 0$, then $A^+ = \Phi$.

Proof: By the Cayley-Hamilton Theorem, AA^* satisfies its characteristic equation so that

$$f(AA^*) = (AA^+)^n + a_1(AA^*)^{n-1} + \ldots + a_k(AA^*)^{n-k} + \ldots$$

$$+ a_{n-1}AA^* + a_n I = \Phi .$$

If $k \neq 0$ is the largest integer such that $a_k \neq 0$, then if we take $B = AA^*$ and $B^0 = I$, then

$$B^n + a_1B^{n-1} + \ldots + a_kB^{n-k} = \Phi .$$

$$B^{n-k}(B^k + a_1B^{k-1} + \ldots + a_kI) = \Phi .$$

This latter equation proves the existence of a solution to the equation

$$B^{n-k}X =$$

and, the general solution is given by

$$X = B^{n-1} + Y - (B^{n-k})^+ \ B^{n-k}Y$$

$$= Y - (B^{n-k})^+ \ B^{n-k}Y \ .$$

In particular there exists a Y such that

$$B^k + a_1B^{k-1} + \ . \ . \ . + a_kI = Y \ 0 \ (B^{n-k})^+B^{n-k}Y \ .$$

Since $B = AA^*$ is normal then for each integer p, $(B^P)^+ = (B^+)^P$ and $B^+B = BB^+$. This fact together with the fact that B^+B is idempotent implies

$$(B^{n-k})^+B^{n-k} = (B^+)^{n-k}B^{n-k} = (B^+B)^{n-k} = B^+B.$$

Hence,

$$B^k + a_1B^{k-1} + \ . \ . \ . + a_kI = Y_1 - B^+BY_1.$$

But $(AA^*)^+ \ (AA^*) = AA^+$ so that

$$B^+B = (AA^*)^+ \ (AA^*) = AA^+.$$

Therefore,

$$B^k + a_1B^{k-1} + \ . \ . \ . + a_kI = Y_1 - AA^+Y_1.$$

Multiplying on the left by A^+ we get

$$A^+B^k + a_1A^+B^{k-1} + \ldots + a_kA^+ = \Phi .$$

Noting that $A^+AA^* = A^+B = A^*$ we obtain

$$A^*B^{k-1} + a_1A^*B^{k-2} + \ldots + a_{k-1}A^* = -a_kA^+$$

or

$$A^+ = -a_k^{-1}A^* \; [(AA^*)^{k-1} + a_1(AA^*)^{k-2} + \ldots + a_{k-1}I].$$

If $k = 0$, then $(AA^*)^n = \Phi$ hence $A = \Phi$ and $A^+ = A^T = \Phi$.

As a consequence of Theorem 6.1, D. K.Faddeev's modification of Leverrier's method [39] can be further modified to describe a computing algorithm for the generalized inverse of A. Consider the following sequence $A_o, A_1, A_2, \ldots, A_k$:

$$\begin{array}{lll}
A_o = \Phi & -1 = q_o & B_o = I \\
A_1 = AA^* & \text{tr } A_1 = q_1 & B_1 = A_1 - q_1 I \\
A_2 = AA^*B_1 & \dfrac{\text{tr } A_2}{2} = q_2 & B_2 = A_2 - q_2 I \\
\vdots & \vdots & \vdots \\
A_{k-1} = AA^*B_{k-2} & \dfrac{\text{tr } A_{k-1}}{k-1} = q_{k-1} & B_{k-1} = A - qI \\
A_k AA^*B_{k-1} & \dfrac{\text{tr } A_k}{k} = q_k & B_k = A_k - q_k I
\end{array}$$

Faddeev shows that $q_k = a_i$, $i = 1, \ldots, k$; hence by Theorem 6.1 we have either that $A^+ = \Phi$ or

$$A^{+} = -a_k^{-1}A^{*}[(AA^{*})^{k-1} + a_1(AA^{*})^{k-2} + \ldots + a_{k-1}I] =$$

$$= -a_k^{-1}A^{*}B_{k-1} .$$

Example: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Then

$$AA^{*} = \begin{pmatrix} 3 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix} .$$

It is computational to confirm that $f(\lambda) = (-1)(\lambda^3 - 8\lambda^2 + 15\lambda)$, hence that $k = 2$, $a_2 = 15$, $a_1 = -8$, and $a_0 = 1$. Thus

$$A^{+} = -\frac{1}{15} A^{*}(AA^{*} - 8I)$$

$$= \frac{1}{15}\begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & -5 & -1 \\ 3 & 0 & 3 \end{pmatrix} .$$

It should be noted in using this algorithm that if A is n by m with $m > n$, then one could replace A with A^{*} in the above and using the relationship $(A^{*})^{+} = (A^{+})^{*}$ obtain A^{+}. This enables one to work with the smallest possible matrices in the characteristic polynomial.

Method II (Decell)

Decell [32] has presented another explicit form which gives rise to an algorithm for the computation of the pseudo inverse. This particular form arises from the following theorems.

Theorem 6.2: For any matrix A, $A = WAY$, where W and Y are, respectively, any solutions of

$$WAA^* = A^* \tag{1}$$

$$A^*AY = A^* . \tag{2}$$

Proof: Properties P3 and P4 of Theorem 3.1 guarantee a solution to Equations 1 and 2, in particular $W = Y = A^+$. If W and Y are solutions then

$$AWAA^* = AA^* \quad \text{and} \quad A^*AYA = A^*A.$$

But since $BAA^* = CAA^*$ implies $BA = CA$, then

$$AWA = A \quad \text{and} \quad AYA = A.$$

Also,

$$WAA^*W^* = A^*W^* \quad \text{and} \quad Y^*A^*YA = Y^*A^*$$

so that

$$(WA)^* = WA \quad \text{and} \quad (AY)^* = AY.$$

Therefore, $X = WAY$ satisfies the Penrose equations:

$$A(WAY)A = AYA = A$$

$$(WAY)A(WAY) = WAWAY = WAY$$

$$[(WAY)A]^* = A^*AYW^* = A^*W^* = (WA)^* = WAYA$$

$$[A(WAY)]^+ = Y^*WAA^* = Y^*A^* = (AY)^* = AWAY .$$

Thus $A^+ = X = WAY$.

Corollary 6.1: For any matrix A, $A^+ = A^*S_1AS_2A^*$ where S_1 and S_2 are, respectively, any solution of

$$(AA^*)S_1(AA^*) = (AA^*) \tag{3}$$

and

$$(A^*A)S_2(A^*A) = (A^*A) . \tag{4}$$

Proof: Since $BAA^* = CAA^*$ implies $BA = CA$, taking transposes we get

$$AA^*B^* = AA^*C^* \quad \text{implies} \quad A^*B^* = A^*C^* .$$

Applying this result to Equation 3

$$AA^*S_1AA^* = AA^* \quad \text{implies} \quad A^*S_1AA^* = A^*$$

so that $A^*S_1 = W$ satisfies Equation 1 of Theorem 6.2. Similarly, if $A^+AS_2A^*A = A^*A$, then

$$A^*AS_2A^* = A^* ,$$

hence $Y = S_2A^*$ satisfies Equation 2 of Theorem 6.2. Then by the conclusion of this theorem we have the result, namely

$$A^+ = A^*S_1AS_2A^* .$$

<u>Theorem 6.3</u>: <u>If</u> B <u>is a matrix and there exist nonsingular matrices</u> P <u>and</u> Q <u>such that</u> $PBQ = E$ <u>is idempotent then</u> $\overline{B} = QEP$ <u>is a solution of</u> $BXB = B$.

<u>Proof</u>: Since

$$B = P^{-1}EQ^{-1} ,$$

then

$$B\overline{B}B = (P^{-1}EQ^{-1})QEP(P^{-1}EQ^{-1}) = P^{-1}E^3Q^{-1} = P^{-1}EQ^{-1} = B .$$

Corollary 6.1 and Theorem 6.3 suggest an algorithm for computing the pseudoinverse of a complex matrix F. Recalling that

$$F^+ = (F^*F)^+F^*,$$

we reduce the problem of finding F^+ to that of finding the pseudo-inverse of the hermitian matrix $F^*F = C$. Since $(C^2)^* = C$, there exist nonsingular matrices P and Q such that

$$PC^2Q = \begin{pmatrix} I_r & \Phi \\ \Phi & \Phi \end{pmatrix} = I_o$$

where I_r is a rank r identity matrix. We let $C = A$ in Corollary 6.1, so that

$$A^*A = AA^* = C^*C = CC^* = C^2 .$$

Then according to Theorem 6.3, choose solutions $S_1 = S_2 = QI_oP$ so that

$$C^{+} = (CX_1)^2 C ,$$

$$(F^{*}F)^{+} = C^{+} ,$$

and

$$F^{+} = C^{+}F^{*} .$$

Computing programs are available for calculating S_1 and S_2 (e.g., STORM, Statistically Oriented Matrix Program, IBM). In general, these programs only compute some solution of the equation $AXA = A$, usually different from A^{+}. These results allow one to construct a solution to all four Penrose equations given only a solution to the first, namely, $AXA = A$.

6.5 Greville Method

Greville presents a concise recursive algorithm for computing the pseudo inverse of a matrix. The algorithm is as follows:

Let a_k denote the k^{th} column of a given matrix A, and let A_k denote the matrix consisting of the first k columns. Consider A_k in the partitioned form (A_{k-1}, a_k).

Compute $d_k = A^{+}_{k-1}\, a_k$

and $C_k = a_k - A_{k-1}\, d_k$. If $C_k \neq 0$, let $b_k = C^{+}_k$.

If $C_k = 0$, compute $b_k = (1 + d^T_k\, d_k)^{-1}\, d^T_k\, A^{+}_{k-1}$.

Then

$$A^{+}_k = \begin{pmatrix} A^{+}_{k-1} - d_k\, b_k \\ b_k \end{pmatrix}.$$

To initiate the process, take $A_1^+ = 0$ if a_1 is a zero vector; otherwise, $A_1^+ = (a_1^T\, a_1)^{-1/2}\, a_1^T$.

This algorithm is very easy to follow and compute. It requires only one decision in each cycle. It requires no inverse of a matrix.

<u>Example</u>: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \text{then } A_1^+ = (1/2,\ 0,\ 1/2).$$

$$d_2 = A_1^+ a_2 = (1/2,\ 0,\ 1/2) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = (1/2)$$

$$c_2 = a_2 - A_1 d_2 = \begin{pmatrix} -1/2 \\ 1 \\ 1/2 \end{pmatrix} \neq 0$$

Hence $b_2 = c_2^+ = (-1/3,\ 2/3,\ 1/3)$ and $A_2^+ = \begin{pmatrix} A_1^+ - d_2 b_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} 2/3, & -1/3, & 1/3 \\ -1/3, & 2/3, & 1/3 \end{pmatrix}$

Now $d_3 = A_2^+ a_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$$c_3 = a_3 - A_2\, d_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \Phi$$

Hence
$$b_3 = (1 + d_3^+ d_3)^{-1}\, d_3^T\, A_2^+$$
$$= (1 + 2)^{-1}\, (1,\ -1) \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \end{pmatrix}$$
$$= (1/3,\ \ -1/3,\ 0)$$

Thus $$A_3^+ = \begin{pmatrix} A_2^+ - d_3 b_3 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1/3 & 0 & 1/3 \\ 0 & 1/3 & 1/3 \\ 1/3 & -1/3 & 0 \end{pmatrix}$$

Now $d_4 = \begin{pmatrix} 2/3 \\ 1/3 \\ 1/3 \end{pmatrix}$ and $c_4 = 0$ so that $b_4 = (1/5,\ 0,\ 1/5)$

and $$A_4^+ = \begin{pmatrix} A_3^+ - d_4 b_4 \\ b_4 \end{pmatrix} = \frac{1}{15} \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & -5 & -1 \\ 3 & 0 & 3 \end{pmatrix} .$$

To show that this method is programmable for computer solution, a flow chart is included.

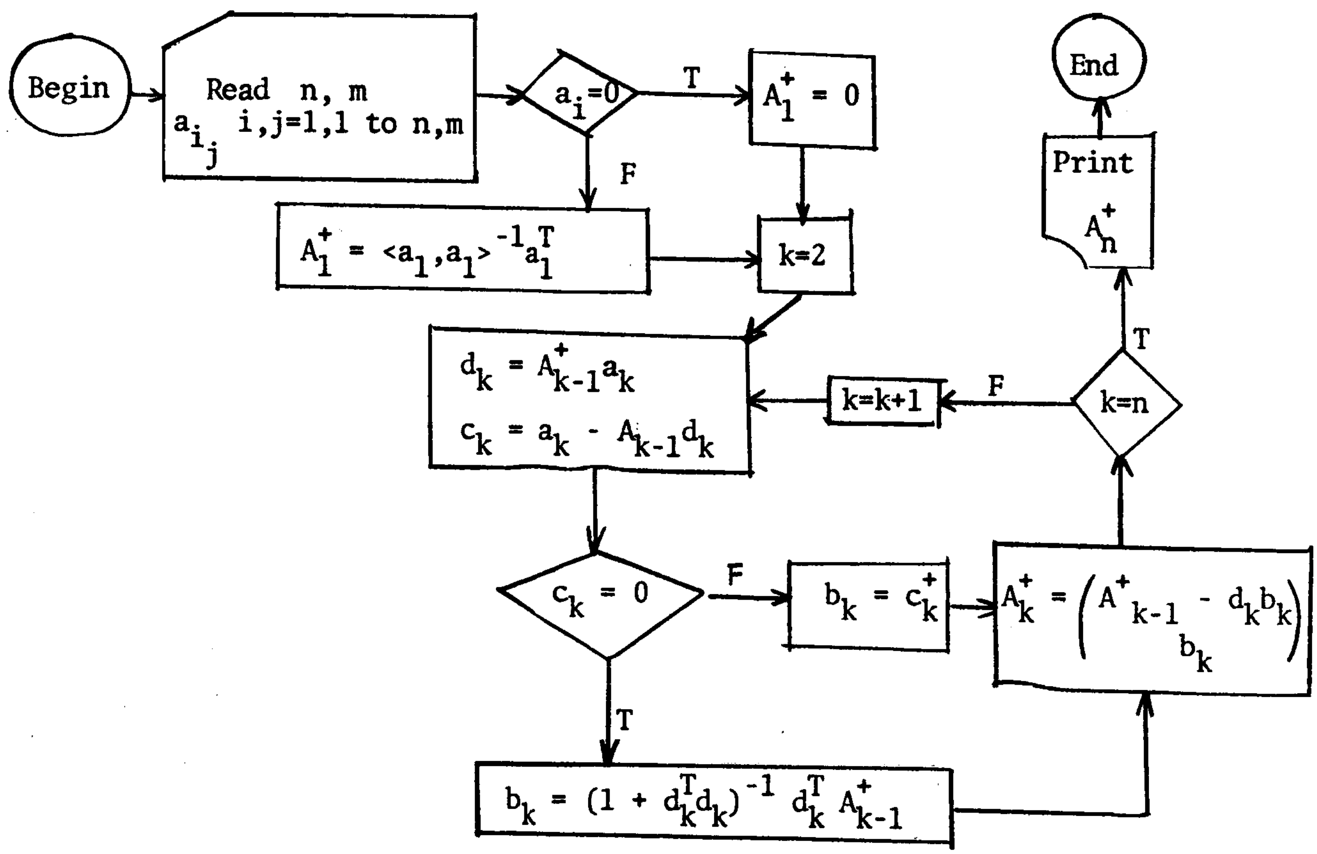

Fig. 2 Flow Chart for Greville Method

6.6 Hestenes Method

Hestenes obtained a scheme for inverting matrices by a process called biorthogonalization. This process can be extended and modified to all rectangular matrices of any rank.

The concept of biorthogonality can be expressed in matrix form as follows. The vectors (in an n-dimensional space) $u_1, u_2, \ldots,$ u_n can be considered to be the column vectors of a matrix U and v_1, 1 1 1, v_n, the row vectors of a matrix V. The set is biorthogonal if

$$VU = I .$$

If $n = m$, then V is the inverse of U.

Hestenes poses and solves the problem: Given two sets $u_1, \ldots,$ u_n and $v_1, \ldots, v_n$ of n vectors in an m-dimensional space $(m \geq n)$, obtain a biorthogonal system by modifying the v's. The solution is arrived at by letting $v_1^{(0)}, \ldots, v_n^{(0)}$ be the initial choice for the v's. These vectors are modified successively in n steps. After n steps the vectors $v_1^{(n)}, \ldots, v_n^{(n)}$ will be a solution to the problem.

In the k^{th} step the vectors $v_k^{(i-1)}$ are transformed into a new set $v_i^{(k)}$ by the following computations.

$$c_{kk} = \left\langle v_k^{(k-1)}, u_k \right\rangle \quad \text{where} \quad \langle a,b \rangle = (a^T b)^{1/2} .$$

$$c_k = c_{kk}^{-1}$$

$$c_{jk} = \left\langle v_j^{(k-1)}, u_k \right\rangle \quad j \neq k$$

$$v^{(k)} = c^{(k)} v^{(k-1)}$$

where $c_{ij}^{(k)} = \delta_{ij}$ $(j \neq k)$, $c_{kk}^{(k)} = c_k$, $c_{ij}^{(k)} = -c_{ik}c_k$ for $i \neq k$.

The only difficulty is to insure that $c_{kk} \neq 0$. This will not arise if $V^{(o)}$ is taken to be U^*.

To compute the pseudo inverse of a matrix, the method is modified by adding rows to the original matrix which are orthogonal and which raise the rank of the matrix to its column dimension. Then the method is applied to the resulting matrix. The pseudo inverse of the original matrix is obtained by deleting the last added columns of $V^{(n)}$.

Example: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Add two rows to A which are orthogonal to all other rows.

$$U = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ \hdashline 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$$

and let

$$V^{(0)} = U^* = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 \end{pmatrix}$$

Then $c_{11} = \left\langle v_1^{(0)}, u_1 \right\rangle = (1, 0, 1, 1, 0)(1, 0, 1, 1, 0)^T = 3$

so that $c_1 = 1/3$.

$$c_{21} = \left\langle v_2^{(0)},\ u_1 \right\rangle = (0,\ 1,\ 1,\ 0,\ 1)(1,\ 0,\ 1,\ 1,\ 0)^T = 1,$$

$$c_{21}^{(1)} \ = \ -c_{21}\ c_1 = -1/3$$

$$c_{31} = \left\langle v_3^{(o)},\ u_1 \right\rangle = (1,\ -1,\ 0,\ 0,\ 1)(1,\ 0,\ 1,\ 1,\ 0)^T = 1,$$

$$c_{31}^{(1)} \ = -c_{31}\ c_1 = -1/3$$

$$c_{41} = \left\langle v_4^{(o)},\ u_1 \right\rangle = (1,\ 0,\ 1,\ -1,\ -1)(1,\ 0,\ 1,\ 1,\ 0)^T = 1,$$

$$c_{41}^{(o)} \ = -c_{41}\ c_1 = -1/3$$

Hence

$$c^{(1)} = \begin{pmatrix} 1/3 & 0 & 0 & 0 \\ -1/3 & 1 & 0 & 0 \\ -1/3 & 0 & 1 & 0 \\ -1/3 & 0 & 0 & 1 \end{pmatrix}$$

and

$$V_1^{(1)} = \begin{pmatrix} 1/3 & 0 & 0 & 0 \\ -1/3 & 1 & 0 & 0 \\ -1/3 & 0 & 1 & 0 \\ -1/3 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1/3 & 0 & 1/3 & 1/3 & 0 \\ -1/3 & 1 & 2/3 & -1/3 & 1 \\ 2/3 & -1 & -1/3 & -1/3 & 1 \\ 2\text{-}3 & 0 & 2/3 & -4/3 & -1 \end{pmatrix}$$

Now

$$c_{22} = \left\langle v_2^{(1)},\ u_2 \right\rangle = (-1/3,\ 2/3,\ -1/3,\ 1)(0,\ 1,\ 1,\ 0,\ 1)^T = 8/3,$$

$$c_{22}^{(2)} \ = \ 3/8$$

$$c_{12} = \left\langle v_1^{(1)},\ u_2 \right\rangle = (1/3,\ 0,\ 1/3,\ 1/3,\ 0)(0,\ 1,\ 1,\ 0,\ 1)^T = 1/3,$$

$$c_{12}^{(2)} \ = \ -1/8$$

$$c_{32} = \left\langle v_3^{(1)},\ u_2 \right\rangle = (2/3,\ -1,\ -1/3,\ -1/3,\ 1)(0,\ 1,\ 1,\ 0,\ 1)^T = -1/3,$$

$$c_{32}^{(2)} \ = \ 1/8$$

$$c_{42} = v_4^{(1)}, u_2 = (2/3,\ 0,\ 2/3,\ -4/3,\ -1)(0,\ 1,\ 1,\ 0,\ 1)^T = -1/3,$$

$$c_{42}^{(2)} = 1/8.$$

Hence

$$V^{(2)} = C^{(2)}V^{(1)} = \begin{pmatrix} 1 & -1/8 & 0 & 0 \\ 0 & 3/8 & 0 & 0 \\ 0 & 1/8 & 1 & 0 \\ 0 & 1/8 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/3 & 0 & 1/3 & 1/3 & 0 \\ -1/3 & 1 & 2/3 & -1/3 & 1 \\ 2/3 & -1 & -1/3 & -1/3 & 1 \\ 2/3 & 0 & 2/3 & -4/3 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 3/8 & -1/8 & 2/8 & 3/8 & -1/8 \\ -1/8 & 3/8 & 2/8 & -1/8 & 3/8 \\ 5/8 & -7/8 & -2/8 & -3/8 & 9/8 \\ 5/8 & 1/8 & 6/8 & -11/8 & -7/8 \end{pmatrix}.$$

Likewise

$$V^{(3)} = \begin{pmatrix} 6/21 & 0 & 6/21 & 9/21 & -6/21 \\ -2/21 & 7/21 & 5/21 & -3/21 & 9/21 \\ 5/21 & -7/21 & -2/21 & -3/21 & 9/21 \\ 15/21 & 0 & 15/21 & -30/21 & -15/21 \end{pmatrix}$$

and

$$V^{(4)} = \begin{pmatrix} 1/5 & 0 & 1/5 & 3/5 & -1/5 \\ -1/15 & 1/3 & 4/15 & -1/5 & 2/5 \\ 4/15 & -1/3 & -1/15 & -1/5 & 2/5 \\ 1/5 & 0 & 1/5 & -2/5 & -6/5 \end{pmatrix}$$

Hence the pseudo inverse of A is obtained by deleting the last two columns of $V^{(4)}$.

$$A^+ = \begin{pmatrix} 1/5 & 0 & 1/5 \\ -1/15 & 1/3 & 4/15 \\ 4/15 & -1/3 & -1/15 \\ 1/5 & 0 & 1/5 \end{pmatrix}.$$

Hestenes makes the statement that the process requires n divisions, $2n^3$ multiplications, and $2n^2(n - 1)$ additions.

6.7 Pyle Method

It has been shown that the best approximate solution of the equation $Ax = b$ is unique and is given by $x = A^+b$. Pyle has facilitated applications of this type by contributing an algorithm for obtaining A^+b, assuming $AA^+b = b$.

This computational technique is a variation of the gradient projection method. The basic steps in the procedure require the application of the Gram-Schmidt orthogonalization process first to the column vectors of A^*, then, if A is not full row rank, to the column vectors of A. A computer program (Fortran IV for the IBM 7090) using the method has been tested and used in connection with the generalized inverse-eigenvector method for solving problems in linear programming.

Recall the following properties: A necessary and sufficient condition that $Ax = b$ be solvable is that $AA^+b = b$, and if this condition is met, then $x = A^+b = (I - A^+A)y$ for some (compatible) y.

Also, A, A^*A, A^+, and A^+A all have rank equal to trace A^+A, $(A^+)^* = (A^*)^+$ so that we may assume that in $A(m \times n)$ that $m \leq n$.

Assume the $m \times n$ complex matrix A satisfies $AA^+b = b$. Let the row vectors of A be denoted by $\{a^{(i)}\}$, $(i = 1, \ldots, m)$; the elements of b by $\{b^{(i)}\}$, $(i = 1, \ldots, m)$, and the i^{th} equation in $Ax = b$ may be written either as the <u>vector product</u> $a^{(i)}x = b$, $(i = 1, \ldots, m)$, or the <u>inner product</u> $(x, a^{(i)*}) = b_i$, where the inner product (u, v) of two vectors is defined by

$(u, v) = \sum_{i=1}^{n} u_i \bar{v}_i$. The first r linearly independent $a^{(i)*}$ will be designated $\{\tilde{a}^{(i)*}\}$, $(i = 1, \ldots, r)$. The $\{\tilde{a}^{(i)*}\}$ are determined by applying the Gram-Schmidt orthogonalization process to the vectors $\{a^{(i)*}\}$, $(i = 1, \ldots, m)$, obtaining the orthonormal system of vectors $\{\eta^{(i)}\}$, $(i = 1, \ldots, r)$, where

$$\eta^{(1)} = \frac{1}{||\tilde{a}^{(i)*}||} \tilde{a}^{(i)*} , \quad \eta^{(k)} = \frac{1}{||\tilde{\eta}^{(k)}||} \tilde{\eta}^{(k)} , \quad k = 2, 3, \ldots, r$$

with

$$\tilde{\eta}^{(k)} = \tilde{a}^{(k)*} - \sum_{i=1}^{k-1} (\tilde{a}^{(k)*} , \eta^{(i)}) \eta^{(i)} .$$

(Remarks: $\tilde{a}^{(i)*}$ is the first non-zero $a^{(i)*}$. The notation $||u||$, the length of u, means $\sqrt{(u, u)}$ for any vector u.)

Determine successively the scalar values $\{a_k\}$ and vectors $x^{(k)}$ for $(k = 1, \ldots, r)$ by using the relations

$$\alpha_k = \frac{\tilde{b}_k - (x^{(k-1)} , \tilde{a}^{(k)*})}{(\eta^{(k)} , \tilde{a}^{(k)*})}$$

and

$$x^{(k)} = x^{(k-1)} + \alpha_k \eta^{(k)}$$

where $x^{(o)} = \Theta$, (the null vector) and $\{\tilde{b}_k\}$, $(i = 1, \ldots, r)$

consists of those elements of $\{ b_i \}$, $(i=1, \ldots, m)$ corresponding to the linearly independent subset $\{ a^{(i)*} \}$ determined in the application of the Gram-Schmidt process, listed in the same order.

Theorem 6.4: $x^{(r)} = A^{+}b.$

Proof: The proof is accomplished by showing by induction that $Ax^{(r)} = b$, then noting the consistency assumption $AA^{+}b = b$. Since $x^{(1)} = \alpha_1 \eta^{(1)}$, where

$$\eta^{(1)} = \frac{1}{||\tilde{a}^{(1)*}||} \tilde{a}^{(1)*} \quad \text{and} \quad \alpha_1 = \frac{\tilde{b}_1}{(\eta^{(1)}, \tilde{a}^{(1)*})} ,$$

then

$$\tilde{a}^{(1)}x^{(1)} = \tilde{a}^{(1)} \frac{1}{||\tilde{a}^{(1)*}||} \tilde{a}^{(1)*} \frac{\tilde{b}_1}{(\eta^{(1)}, \tilde{a}^{(1)*})} =$$

$$= [\frac{\tilde{b}_1 ||a^{(1)*}||}{(\tilde{a}^{(1)*}, \tilde{a}^{(1)*})||\tilde{a}^{(1)*}||}] \tilde{a}^{(1)}\tilde{a}^{(1)*} = \tilde{b}_1 .$$

By the Gram-Schmidt process $M(\tilde{a}^{(1)*}, \ldots, \tilde{a}^{(k)*}) = M(\eta^{(1)}, \ldots, \eta^{(k)})$ where $M(\eta^{(1)}, \ldots, \eta^{(k)})$ is the linear manifold spanned by the vectors $\{\eta^{(i)}\}$, $(i = 1, \ldots, k)$. Suppose $x^{(k-1)}$ is a solution of the $k-1$ equations of the system $Ax = b$ corresponding to $\{\tilde{a}^{(1)*}\}$, $(i = 1, \ldots, k-1)$, i.e., that $\tilde{a}^{(i)} x^{(k-1)} = \tilde{b}_i$, $(i = 1, \ldots, k-1)$. We must show that

$\tilde{a}^{(i)}\, x^{(k)} = b_i$, $(i = 1, \ldots, k)$. Now, $(\eta^{(k)}, \eta^{(i)}) = 0$ for $(i = 1, \ldots, k-1)$, together with $M(\eta^{(1)}, \ldots, \eta^{(k-1)}) = M(\tilde{a}^{(1)*}, \ldots, \tilde{a}^{(k-1)*})$ implies

$$\tilde{a}^{(i)}\ {}^{(k)} = (\eta^{(k)}, \tilde{a}^{(i)*}) = 0$$

and thus

$$\tilde{a}^{(i)}x^{(k)} = \tilde{a}^{(i)}(x^{(k-1)} + \alpha_k \cdot \eta^{(i)}) = \tilde{a}^{(i)}x^{(i-1)} + 0 = \tilde{b}_i$$

for $(i = 1, \ldots, k-1)$.

Now if

$$\alpha_k = \frac{\tilde{b}_k - (x^{(k-1)}, \tilde{a}^{(k)*})}{(\eta^{(k)}, \tilde{a}^{(k)*})},$$

then

$$\tilde{a}^{(i)}x^{(k)} = \tilde{a}^{(k)}x^{(k-1)} + \left[\frac{\tilde{b}_k - (x^{(k-1)}, \tilde{a}^{(k)*})}{(\eta^{(k)}, \tilde{a}^{(k)*})}\right] \tilde{a}^{(k)}\eta^{(k)} = \tilde{b}_k .$$

Hence,

$$Ax^{(r)} = b .$$

Under the consistency assumption $AA^{+}b = b$, the equations corresponding to any linearly dependent columns of A^{*} will be automatically satisfied by $x^{(r)}$, or the system is inconsistent. In practice, if dependent columns of A^{*} are encountered, the corresponding equations may be checked for consistency by substitution of the current $x^{(k)}$ and then set aside since they do not enter into the continued application of the algorithm.

Thus $Ax^{(r)} = b$ where $x^{(r)}$ is a linear combination of columns of A^*. Since

$$(I - A^+A)A^* = A^* - A^+AA^* = A^* - (A^+A)^*A^*$$

$$A^* - (AA^+A)^* = A^* - A^* = \Phi$$

then $x^{(r)}$ is an eigenvector of $(I - A^+A)$ for eigenvalue $\lambda = 0$. That is $A^+Ax^{(r)} = x^{(r)}$. By Property 3.25, $x^{(r)} = A^+b + (I - A^+A)y$ for some y.

Thus

$$x^{(r)} = A^+Ax^{(r)} = A^+A[A^+b + (I - A^+A)y]$$

$$= A^+AA^+b + (A^+A - A^+AA^+A)y$$

$$= A^+b + (A^+A - A^+A)y$$

$$= A^+b \ .$$

Now, if $AA^+I = AA^+ = I$, the above algorithm may be used to obtain, successively, the columns of A^+ by taking successive b vectors equal to the columns of the $m \times n$ identity matrix. Unfortunately $AA^+ = I$ if and only if $r = m$, a somewhat special case. Usually $AA^+ \neq I$ and the method must be extended as indicated by the following theorems.

Corollary 6.2: $AX = AA^+$ is always solvable and application of the gradient projection algorithm with b vectors chosen, successively, as columns of AA^+, yields $X = A^+AA^+ = A^+$.

Proof: $A^+AA^+ = A^+$ implies $AA^+(AA^+) = AA^+$ for any matrix A. But this is the consistency hypothesis.

Corollary 6.2 would be of little interest except for the following theorem which gives a method for obtaining AA^+.

Theorem 6.5: *Let* $\{\xi^{(i)}\}$, $(i = 1, \ldots, r)$ *be the orthonormal set of vectors obtained by applying the Gram-Schmidt process to the columns of* A. *Then*

$$AA^+ = \sum_{i=1}^{r} \xi^{(i)} \xi^{(i)*} \quad .$$

Proof: The vectors $\{\xi^{(i)}\}$, $(i = 1, \ldots, r)$ provide an orthonormal basis for the column space of A since they are orthonormal linear combinations of the linearly independent columns of A. Since $AA^+A = A$, the $\{\xi^{(i)}\}$, $(i = 1, \ldots, r)$ are eigenvectors of AA^+ corresponding to the eigenvalue $\lambda = 1$. Since rank A^+ = rank A^+A = trace A^+A, and since $(A^+)^+ = A$, then rank A = rank AA^+. Therefore, r = rank A = rank AA^+ = dimension of the range of AA^+, (written $R(AA^+)$). Thus the $\{\xi^{(i)}\}$, $(i = 1, \ldots, r)$ provide an orthonormal basis for $R(AA^+)$. Extending the $\{\xi^{(i)}\}$, $(i = 1, \ldots, r)$ to an orthonormal basis $\{\xi^{(i)}\}$, $(i = 1, \ldots, m)$ of C^m, m-dimensional complex unitary space, yields vectors $\{\xi^{(i)}\}$, $(i = r + 1, \ldots, m)$ which are eigenvectors of AA^+ for eigenvalue $\lambda = 0$, since for $(i = r + 1, \ldots, m)$, $\Phi = A^*\xi^{(i)}$ and thus

$$\Phi = (A^+)^*A^*\xi^{(i)} = (AA^+)^* \xi^{(i)} = AA^+ \xi^{(i)} \; .$$

Consider $T = [\xi^{(1)}, \ldots, \xi^{(r)}]$, the matrix whose columns are the $\{\xi^{(i)}\}$, $(i = 1, \ldots, r)$. Then $AA^+T = T$ implies $AA^+TT^* = TT^*$. It is easily verified that

$$\sum_{i=1}^{m} \xi^{(i)}\xi^{(i)*} = I$$

hence

$$I - \sum_{i=r+1}^{m} \xi^{(i)}\xi^{(i)*} = \sum_{i=1}^{r} \xi^{(i)}\xi^{(i)*} = TT^* .$$

Thus

$$TT^* = AA^+TT^* = AA^+(I - \sum_{i=r+1}^{m} \xi^{(i)}\xi^{(i)*}) = AA^+ + 0 = AA^+$$

since the $\{\xi^{(i)}\}$, $(i = r + 1, \ldots, m)$ are eigenvectors of AA^+ corresponding to eigenvalue $\lambda = 0$.

Example: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} .$$

Thus

$$a^{(1)*} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} , \quad a^{(2)*} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} , \quad a^{(3)*} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Apply the Gram-Schmidt process, obtaining

$$\eta^{(1)} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\0\\1\\1\end{pmatrix}, \quad \eta^{(2)} = \frac{1}{\sqrt{15}}\begin{pmatrix}1\\3\\-2\\1\end{pmatrix}, \quad \tilde{\eta}^{(3)} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

Thus $\tilde{a}^{(1)*} = a^{(1)*}$, $\tilde{a}^{(2)*} = a^{(2)*}$ and $r = \text{rank } A = 2$. Since $r = 2 < 3 = m$, AA^+ must be computed. Let the column vectors of A be designated

$$c^{(1)} = \begin{pmatrix}1\\0\\1\end{pmatrix}, \quad c^{(2)} = \begin{pmatrix}0\\1\\1\end{pmatrix}, \quad c^{(3)} = \begin{pmatrix}1\\-1\\0\end{pmatrix}, \quad c^{(4)} = \begin{pmatrix}1\\0\\1\end{pmatrix}.$$

Apply the Gram-Schmidt process, obtaining the $r = 2$ vectors

$$\xi^{(1)} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix} \qquad \xi^{(2)} = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\1\end{pmatrix}.$$

Then

$$AA^+ = \sum_{i=1}^{2} \xi^{(i)}\xi^{(i)*} = \frac{1}{3}\begin{pmatrix}-2 & -1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{pmatrix}.$$

Note that trace $AA^+ = 2 = r$.

Now solve $Ax = b^{(i)}$ $(i = 1, 2, 3)$ where

$$b^{(1)} = \frac{1}{3}\begin{pmatrix}2\\-1\\1\end{pmatrix}, \quad b^{(2)} = \frac{1}{3}\begin{pmatrix}1\\-2\\1\end{pmatrix}, \quad b^{(3)} = \frac{1}{3}\begin{pmatrix}1\\1\\2\end{pmatrix}.$$

(Observe that, in general, only r components of each $b^{(i)}$ are required; in this case, the first r.)

Applying the gradient projection algorithm, obtain

corresponding to $b^{(1)}$: $\alpha_1 = \frac{2\sqrt{3}}{9}$, $\alpha_2 = \frac{-\sqrt{15}}{45}$,

$$x^{(1)} = \frac{2}{9}\begin{pmatrix}1\\0\\1\\1\end{pmatrix} ; \quad x^{(2)} = \frac{1}{15}\begin{pmatrix}3\\-1\\4\\3\end{pmatrix} = \text{the first column of } A^+;$$

corresponding to $b^{(2)}$: $\alpha_1 = \frac{-\sqrt{3}}{9}$, $\alpha_2 = \frac{\sqrt{15}}{9}$,

$$x^{(1)} = \frac{-1}{9}\begin{pmatrix}1\\0\\1\\1\end{pmatrix} ; \quad x^{(2)} = \frac{1}{3}\begin{pmatrix}0\\1\\-1\\0\end{pmatrix} = \text{the second column of } A^+;$$

corresponding to $b^{(3)}$: $\alpha_1 = \frac{\sqrt{3}}{9}$, $\alpha_2 = \frac{4\sqrt{15}}{45}$,

$$x^{(1)} = \frac{1}{9}\begin{pmatrix}1\\0\\1\\1\end{pmatrix} ; \quad x^{(2)} = \frac{1}{15}\begin{pmatrix}3\\4\\-1\\3\end{pmatrix} = \text{the third column of } A^+.$$

Thus,

$$A^+ = \frac{1}{15}\begin{pmatrix}3 & 0 & 3\\-1 & 5 & 4\\4 & -5 & -1\\3 & 0 & 3\end{pmatrix} .$$

6.8. Ben-Israel and Werson Method

Ben-Israel and Werson presented a method for computing the pseudo inverse of a given matrix which is equivalent to the methods of sections two and three. However, a condensed tableau is presented to facilitate the computation procedure. It does require the inversion of only one matrix instead of two as in sections two and three.

The problem method is formulated as follows:

Let A be the given matrix. Let E be a nonsingular matrix and P a permutation matrix such that

$$EA^*AP = \begin{pmatrix} I_r \; D \\ \hline 0 \;\; 0 \end{pmatrix} = \begin{pmatrix} H^* \\ \hline 0 \end{pmatrix} \tag{1}$$

where r is the rank of A^*A. Using the relations $PP^* = I$ and $A^*AA^+ = A^*$ we have

$$EA^*APP^*A^+ = EA^* \tag{2}$$

which implies

$$P^*A^+ = (EA^*AP)^+EA^* + Z \tag{3}$$

where Z is in the null space of EA^*AP.

We will now show that $Z = 0$. By (1), the columns of Z lie in the null space of H^*. The latter subspace is the orthogonal complement of $R(H) = R(\,(EA^*AP)^*\,) = R(P^*A^*AE^*)$. Since E is nonsingular and $R(A^*A) = R(A^*) = R(A^+)$, we verify that $R(H) = R(P^*A^+)$. On the other hand $R(H) = R(\,(EA^*AP)^*\,) = R(\,(EA^*AP)^+\,)$.

Therefore $R(P^*A^+) = R(\,(EA^*AP)^+\,) = R(H)$, and Z, whose columns lie in $N(H^*)$, must vanish by (3).

Collecting the above results,

$$P^*A^+ = (EA^*AP)^+EA^* = \begin{pmatrix} H^* \\ 0 \end{pmatrix}^+ EA^* = (H^{*+} \quad 0)EA^*. \tag{4}$$

From (1) and (2) it follows that the last $(n - r)$ rows of EA^* are zero; from the definition of H^* it therefore follows that the matrix H^*EA^* consists of the first r rows of EA^*. Therefore

$$HH^+EA^* = H^{+*}H^*EA^* = (H^{*+} \quad 0)EA^* \tag{5}$$

From (4) and (5) and the fact that P is a permutation matrix it follows that

$$A^+ = PHH^+EA^* \tag{6}$$

Finally, if F is an $n \times (n - r)$ matrix such that

$$N(F^*) = R(H)$$

then it is well known that

$$HH^+ = I_n - FF^*$$

and (6) becomes

$$A^+ = P(I_n - FF^*)EA^* . \tag{7}$$

Given $H^* = (I_r \; D)$, a natural choice for F is

$$F = \begin{pmatrix} D \\ \hline -I_{n-r} \end{pmatrix}$$

An elimination method for computing the pseudo inverse may be based either on (4) or (7). Both equations reduce for nonsingular A^*A to $A^+ = EA^*$, and for nonsingular A to the well-known result $A^{-1} = EA^*$, where E is defined by $EAA^* = I_n$. If the matrix A^*A is singular then the method (4) rewritten as

$$A^+ = P \begin{pmatrix} I_r \\ \hline D^* \end{pmatrix} \left((I_r + DD^*)^{-1} \;\; 0 \right) EA^* \tag{8}$$

required the inversion of the r by r matrix $(I_r + DD^*)$. Similarly, if A^*A is singular, the method (7) rewritten as

$$A^+ = P \left[I - \begin{pmatrix} D \\ \hline -I_{n-r} \end{pmatrix} (I_{n-r} + D^*D)^{-1} \left(D^* \;\; -I_{n-r}\right) \right] EA^* \tag{9}$$

requires the inversion of the $(n - r) \times (n - r)$ matrix $(I_{n-r} + D^*D)$.

Since zero rows (or columns) in A result in corresponding zero columns (or rows) in A^+, an obvious reduction in work can be achieved by working with $\tilde{A}$, a matrix obtained from A by striking all zero rows and columns; computing $\tilde{A}^+$ by either (8) or (9) and inserting zero columns and rows to obtain A^+. Another possible reduction in computations and space is by working with A^*A if $m \geq n$ (A is an $m \times n$ matrix), and with AA^* if $m < n$. The latter case results in A^{*+} which must then be transposed to obtain A^+.

For nonsingular matrices the above methods require more operations than the ordinary inversion methods, due to the formation of A^*A. Thus for the nonsingular case: $m = n = r$, both methods require $(5/2\ n^3 - 2n^2 + n/2)$ multiplications, $(3/2\ n^2 - n/2)$ divisions and $(5/2\ n^3 - 2n^2 + n/2)$ additions.

Because the last $(n - r)$ rows of EA^* are zero, in method (9) one need not compute the last $(n - r)$ columns of the matrix $(I - FF^*)$. As in other elemination methods, the above methods depend critically on the correct determination of the rank, which in turn depends on the approximation and roundoff errors.

Example: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

$$A^*A = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix}$$

$$(A^*A \vdots A^*) = \begin{pmatrix} 2 & 1 & 1 & 2 & 1 & 0 & 1 \\ 1 & 2 & -1 & 1 & 0 & 1 & 1 \\ 1 & -1 & 2 & 1 & 1 & -1 & 0 \\ 2 & 1 & 1 & 2 & 1 & 0 & 1 \end{pmatrix} \rightarrow$$

$$\begin{pmatrix} 2 & 0 & 1 & 2 & 1 & 0 & 1 \\ 0 & 3/2 & -3/2 & 0 & -1/2 & 1 & 1/2 \\ 0 & -3/2 & 3/2 & 0 & 1/2 & -1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 1 & 1 & 2 & 1 & 0 & 1 \\ 0 & 3/2 & -3/2 & 0 & -1/2 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow$$

$$\begin{pmatrix} 2 & 1 & 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & -1 & 0 & -1/3 & 2/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 & 2 & 2 & 4/3 & -2/3 & 1/3 \\ 0 & 1 & -1 & 0 & -1/3 & 2/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow$$

$$\left(\begin{array}{cccc:ccc} 1 & 0 & 1 & 1 & 2/3 & -1/3 & 1/3 \\ 0 & 1 & -1 & 0 & -1/3 & 2/3 & 1/3 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right). \quad \text{Hence } r = 2$$

$$D = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, \quad EA^* = \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \end{pmatrix}$$

and $I_2 + DD^* = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$. Thus $(I_2 + DD^*)^{-1} = \begin{pmatrix} 2/5 & 1/5 \\ 1/5 & 3/4 \end{pmatrix}$

and $A^+ = \begin{pmatrix} I_r \\ \hdashline D^* \end{pmatrix} \left((I_r + DD^*)^{-1} \quad 0 \right) \begin{pmatrix} EA^* \\ 0 \end{pmatrix}$

$$= \frac{1}{15} \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & 5 & -1 \\ 3 & 0 & 3 \end{pmatrix} .$$

It should be noted that the matrix to be inverted, $(I_2 + DD^+)$, is the same as the matrix $(AA^* + BB^*)$ of section three which is the same as R^*R of section two. This method has the advantage in that it requires the inversion of only one matrix and has a concise computational layout.

6.9. Ben-Israel and Charnes Method

In their comprehensive paper on generalized inverses, Ben-Israel and Charnes include several alternative expressions for the pseudo inverse of a matrix. Included among those is the Lagrange-Sylvester interpolation polynomial for A^+ .

For any square matrix A, let $\sigma(A)$ denote the spectrum of A (the set of all eigenvalues of A).

Then

$$A^+ = \sum_{\lambda \varepsilon \sigma(A^*A)} \lambda^+ \left(\frac{\prod_{\lambda \neq \theta \varepsilon \sigma (A^*A)} (A^*A- I)}{\prod_{\lambda \neq \theta \varepsilon \sigma (A^*A)} \lambda \ \theta)} \right) A^*$$

where for the real number λ, $\lambda^+ = 1/\lambda$ if $\lambda = 0$ and $\lambda^+ = 0$ if $\lambda = 0$.

In a footnote, it is pointed out by the authors that the Lagrange-Sylvester interpolation polynomial is not a practical way for computing A^+, since it is very sensitive to errors in the computed values of $\sigma(A^*A)$.

Also the computation of the eigenvalues of A^*A is itself a troublesome task although there are schemes for computing them.

<u>Example</u>: Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}, \text{ then } A^*A = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix}$$

and solving $|A^*A - \lambda I| = 0$ for λ we get that $\sigma(A^*A) = \{0, 3, 5\}$.

Hence

$$A^+ = \sum_{\lambda=3,5} \lambda^+ \left(\frac{\prod_{\lambda \neq \theta \varepsilon \{0,3,5\}} (A^*A- \theta I}{\prod_{\lambda \neq \theta \varepsilon \{0,3,5\}} (\lambda-\theta)} \right) A^*$$

$$= 3^+ \frac{(A^*A-0.I)(A^*A-5I)A^*}{(3-0) \quad (3-5)} = 5^+ \frac{(A^*A-0.I)(A^*A-3I)A^*}{(5-0) \quad (5-3)} ,$$

which upon simplifying becomes

$$A^+ = \frac{1}{15}\begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 4 & -5 & -1 \\ 3 & 0 & 3 \end{pmatrix}$$

6.10. Ben-Israel Method

Ben-Israel developed a recursive technique for calculating the pseudo inverse of a matrix which is a generalization of an iterative method, due to Schulz [81], in which the sequence of matrices are defined recursively by

$$x_{n+1} = x_n\,(2I - Ax_n), \quad n = 0, 1, 2, 3, \ldots \tag{1}$$

and is shown to converge to A^{-1} whenever x_o approximates A^{-1}. Ben-Israel pointed out the fact that the computational significance of his iterative method was impaired by the need for knowledge of AA^+. However, in the months between the completion of thses results and their publication, Ben-Israel substantially improved his theorem by finding a starting value which waives the need for AA^+. In the discussion below, we consider Ben=Israel's original iterative method and then his modification of it.

Theorem 6.6: The sequence of matrices defined by

$$X_{n+1} = X_n\,(2P_{R(A)} - AX_n), \quad n = 0, 1, 2, \ldots \tag{2}$$

where X_o is an $n \times m$ complex matrix satisfying

$$X_o = A^* B_o \text{ for some nonsingular } m \times m \text{ matrix } B_o, \tag{3}$$

$$X_o = C_o A^* \text{ for some nonsingular } n \times n \text{ matrix } C_o, \tag{4}$$

$$||AX - P_{R(A)}|| < 1 , \tag{5}$$

$$||XA - P_{R(A^*)}|| < 1 , \tag{6}$$

converges to the pseudoinverse A^+ of A .

<u>Proof</u>: As mentioned before (and as is shown in Ben-Israel and Charnes [5], the pseudoinverse A^+ of A is characterized as the unique solution of the matrix equations

$$AX = P_{R(A)} , \tag{7}$$

$$XA = P_{R(A^*)} . \tag{8}$$

Therefore, it is enough to show that

$$X_{n+1} = X_n(2P_{R(A)} - AX_n)$$

satisfies

$$\lim_{n\to\infty} ||Ax_n - P_{R(A)}|| = 0 , \tag{9}$$

$$\lim_{\to\infty} ||X_n A - P_{R(A^*)}|| = 0 . \tag{10}$$

From Equations 2, 3, and 4, we show that

$$X_n = A^* B_n , \quad n = 0, 1, 2, . . . \tag{11}$$

$$X_n = C_n A^* , \tag{12}$$

where B, C are recursively computed as

$$B_{n+1} - B_n(2P_{R(A)} - AA^*B_n) ,$$

$$C_{n+1} = C_n(2P_{R(A^*)} - A^*AC_n) .$$

Now, from equation 3

$$X_1 = A^*B_o .$$

Suppose

$$X_n = A^*B_n$$

$$= A^*B_{n-1}(2P_{R(A)} - AA^*B_{n-1}) ,$$

then

$$X_{n+1} = X_n(2P_{R(A)} - AX_n)$$

$$= A^*B_n(2P_{R(A)} - AA^*B_n)$$

$$= A^*B_{n+1}$$

The proof that $X_n = C_nA^*$ is similar.

Now, by equation 2 we have

$$X_{n+1} = X_n(2P_{R(A)} - AX_n)$$

$$AX_{n+1} = AX_n(2P_{R(A)} - AX_n)$$

$$AX_{n+1} = AX_n(P_{R(A)} - AX_n) + AX_nP_{R(A)}$$

$$-AX_{n+1} = -AX_nP_{R(A)} - AX_n(P_{R(A)} - AX_n)$$

$$P_{R(A)} - AX_{n+1} = P_{R(A)} - AX_nP_{R(A)} - AX_n(P_{R(A)} - AX_n)$$

$$P_{R(A)} - AX_{n+1} = (P_{R(A)} - AX_n)P_{R(A)} - AX_n(P_{R(A)} - AX_n) ,$$

since $P_{R(A)}$ is idempotent. But by equation 12,

$$X_n = C_n A^*$$

$$AX_n = AC_n A^*$$

$$P_{R(A)} AX_n = AX_n .$$

Also,

$$[AX_n P_{R(A)}]^* = P_{R(A)} X^*_n A^*$$

but

$$X_n^* = AC_n^* ,$$

therefore

$$\begin{aligned} P_{R(A)} X_n^* A^* &= P_{R(A)} AC_n^* A^* \\ &= AC_n^* A^* \\ &= X_n^* A^* \\ &= [AX_n]^* . \end{aligned}$$

It follows that

$$AX_n P_{R(A)} = P_{R(A)} AX_n .$$

Therefore,

$$P_{R(A)} - AX_{n+1} = (P_{R(A)} - AX_n)^2$$

and

$$||P_{R(A)} - AX_{n+1}|| = ||P_{R(A)} - AX_n||^2 , \quad n = 0, 1, 2, . . .$$

which, in view of equation 5, proves equation 9.

To prove equation 10, we write

$$P_{R(A^*)} - X_{n+1}A = P_{R(A^*)} - X_n(2P_{R(A)} - AX_n)A \ ,$$

which is rewritten, by equation 11, as

$$P_{R(A^*)} - X_{n+1}A = P_{R(A^*)} - P_{R(A^*)}X_nA - X_nA + (X_nA)^2$$

since

$$\begin{aligned} X_n(2P_{R(A)})A &= 2X_nA \\ &= X_nA + X_nA \\ &= A^*B_nA + X_nA \\ &= P_{R(A^*)}A^*B_nA + X_nA \\ &= P_{R(A^*)}X_nA + X_nA \ . \end{aligned}$$

But

$$P_{R(A^*)} - P_{R(A^*)}X_nA - X_nA + (X_nA)^2 =$$

$$= P_{R(A^*)}(P_{R(A^*)} - X_nA) - X_nA(P_{R(A^*)} - X_nA) =$$

$$= (P_{R(A^*)} - X_nA)^2 \ ,$$

so that

$$||P_{R(A)} - X_{n+1}A|| \leq ||P_{R(A^*)} - X_nA||^2 \ , \quad n = 0, 1, 2, \ldots$$

which, by equation 6, proves equation 10.

Theorem 6.7: Let A be an arbitrary non-zero complex $m \times n$ matrix of rank r and let

$$\lambda_1(AA^*) \geq \lambda_2(AA^*) \geq \cdots \geq \lambda_r(AA^*)$$

denote the non-zero eigenvalues of AA^*. If the real scalar satisfies

$$0 < \alpha < \frac{2}{\lambda_1(AA^*)} \qquad (13)$$

then the sequence defined by

$$X_o = \alpha A^* \qquad (14)$$

$$X_{k+1} = X_k(2I - AX_k) \ , \quad k = 0, 1, 2, \ldots \qquad (15)$$

converges to A^+ as $k \to \infty$.

Proof: The matrix X_o, defined by equations 14 and 15, satisfies equations 3, 4, 5, and 6. To show that X_o satisfies equation 5

$$||AX_o - P_{R(A)}|| < 1 ,$$

we note that $AA^+ (= P_{R(A)})$ and AA^* are commuting Hermitian matrices with the same range space. The eigenvalues of the $m \times m$ matrix $AX_o - P_{R(A)}$ $(= \alpha AA^* - AA^+)$ are therefore

$$\begin{cases} \alpha\lambda_1(AA^*) - 1 & \text{for} \quad i = 1, 2, \ldots, r \\ 0 & \text{for} \quad i = r+1, \ldots, m \end{cases} \qquad (16)$$

and by equation 13 are all less than 1 in absolute value. That is

$$|\lambda_i(\alpha AA^* - AA^+)| < 1 \ , \quad i = 1, \ldots , m \tag{17}$$

and similarly

$$|\lambda_i(\alpha A^*A - A^+A)| < 1 \ , \quad i = 1, \ldots , n \ . \tag{18}$$

Indeed the non-zero eigenvalues of $(\alpha AA^* - AA^+)$ and $(\alpha A^*A - A^+A)$ are identical. Equations 5 and 6 hold, with Euclidean norm, because of equations 10 and 11, respectively. (Actually, Equations 17 and 18 suffice for the convergence of equation 15.)

Now the process, initiated with $X_o = \alpha A^*$, retains the form of equation 12:

$$X_k = C_k A^*$$

and since

$$A^* P_{R(A)} = A^* \tag{19}$$

it follows that

$$X_k(2P_{R(A)} - AX_k) = X_k(2I - AX_k) \ , \quad k = 0, 1, 2, \ldots \tag{20}$$

and the convergence of equation 15 follows from that of equation 2.

It is interesting to note that it can be shown in a similar manner that the sequence defined by

$$X_{n+1} = (2I - X_k A) X_k \ , \quad k = 0, 1, 2, \ldots$$

with $X_o = A^*$, converges to A^+ .

In applying the method of equation 15, it is not necessary to compute $\lambda_1(AA^*)$. Writing $AA^* = (b_{ij})$, we conclude from the Gershgorin Theorem that

$$\lambda_1(AA^*) \leq \max_{i=1,\ldots,m} \{ \sum_{j=1}^{m} |b_{ij}| \}$$

Therefore equation 13 can be replaced by

$$0 < \alpha < \frac{2}{\max\limits_{i=1,\ldots,m} \{ \sum_{j=1}^{m} |b_{ij}| \}}$$

Examples of this method and application are given in Ben-Israel and Cohen [7].

6.11 Other Methods

In an abstract of a paper presented by E. H. Moore [65] in June, 1920, he gave an explicit formula for each element in the pseudo inverse. This method depends upon the evaluation of many determinants and also presupposes a knowledge of the rank of the matrix A. Its importance is simply its explicit nature. It is impractical for calculating the pseudo inverse.

There are several infinite series representations for A^+ given by Ben-Israel and Charnes [6]. For example,

$$A^+ = \sum_{k=1}^{\infty} A^* (I + AA^*)^{-k}$$

where A^* may not be removed as a factor and the series exist for all A since without loss of generality, null rows or columns may be added to fill out non-square matrices.

Different expressions and computation schemes are continuously appearing in the literature. A thorough analysis of the methods above, however, indicates that some of these methods are quite efficient in obtaining A^+ .

B I B L I O G R A P H Y

1. Barton, C. P. "Pseudoinverses of Rectangular Matrices" Unpublished Master's Thesis, University of Texas (1966).

2. Ben-Israel, A. "An Iterative Method for Generalized Inversion of Matrices," Mathematics of Computation, 91 (1965).

3. Ben-Israel, A. "A Note on an Iterative Method for Generalized Inversion of Matrices," to appear in Mathematics of Computation.

4. Ben-Israel, A., and Charnes, A. "Generalized Inverses and the Bott-Duffin Network Analysis," ONR Research Memo. No. 66. Evanston, Illinois: Northwestern University, The Technological Institute, 1961.

5. Ben-Israel, A., and Charnes, A. "Projection Properties and the Neumann-Euler Expansion for the Moore-Penrose Inverse of an Arbitrary Matrix," ONR Research Memo. No. 40. Evanston, Illinois: Northwestern University, The Technological Institute, 1961.

6. Ben-Israel, A., and Charnes, A. "Contributions to the Theory of Generalized Inverses," Journal of SIAM, 11 (September, 1963), 667-699.

7. Ben-Israel, A., and Cohen, D. "On Iterative Computation of Generalized Inverses and Associated Projections," to appear in Mathematics of Computation.

8. Ben-Israel, A., and Ijiri, Y. "A Report on the Machine Calculation of the Generalized Inverse," Unpublished report. Pittsburgh, Pennsylvania: Carnegie Institute of Technology, 1963.

9. Ben-Israel, A., and Wersan, S. J. "An Elimination Method for Computing the Generalized Inverse of an Arbitrary Complex Matrix," Journal of the Association for Computing Machinery, 10 (October, 1963), 532-537.

10. Berge, Claude "Théorie des Graphes et ses Applications," Dunod, Paris, 1958; translation, "The Theory of Graphs and Its Applications," John Wiley, New York, 1962.

11. Beutler, F. J. "The Operator Theory of the Pseudo-Inverse," Journal of Mathematical Analysis and Applications, 10, 451-493 (1965).

12. Bjerhammar, A. "Application of the Calculus of Matrices to the Method of Least Squares, with Special Reference to Geodetic Calculations," Kungliga Teknisha Högskolans Handlingar (Transactions of the Royal Institute of Technology), Stockholm, Sweden, 49 (1951), 1-68.

13. Bjerhammar, A. "A Generalized Matrix Algebra," Kungliga Tekniska Högskolans Handlingar (Transactions of the Royal Institute of Technology), Stockholm, Sweden, 124 (1958), 1-32.

14. Bjerhammar, A. "Rectangular Reciprocal Matrices with Special Reference to Geodetic Calculations," Bull. Géodésic (1951) 118=220.

15. Boot, J. G. C. "The Computation of the Generalized Inverse of Singular or Rectangular Matrices," American Mathematical Monthly, 70 (March, 1963), 302-303.

16. Bott, R., and Duffin, R. J. "On the Algebra of Networks," American Mathematical Society (January, 1953), 99-109.

17. Boullion, T. L. "Contributions to the Theory of Pseudo-Inverses," Unpublished doctoral dissertation, The University of Texas, 1966.

18. Boullion, T. L., and Odell, P. L. "A Generalization of the Wielandt Inequality," Paper submitted for publication.

19. Bryson, A. E., and Johanson, D. E. "Linear Filtering for Varying Systems Using Measurements Containing Colored Noise" (Unpublished Report).

20. Charnes, A., and Cooper, W. W. "Management Models and Industrial Applications of Linear Programming," John Wiley, New York, 1961.

21. Charnes, A., Cooper, W. W., DeVoe, J. K., and Learner, D. B., Demon: Decision Mapping via Optimum Go-No-Go Networks: A Model for New Products Marketing, Batten, Barton, Durstine, and Osborne, New York, 1963.

22. Charnes, A., Cooper, W. W., and Thompson, G. L. "Constrained Generalized Medians and Linear Programming under Uncertainty," ONR Research Memo. No. 41. Evanston, Illinois, and Pittsburgh, Pennsylvania: Northwestern University, The Technological Institute, and Carnegie Institute of Technology, 1962.

23. Chipman, J. S., and Rao, M. M. "Projections, Generalized Inverses, and Quadratic Forms," Journal of Mathematical Analysis and Applications, 9 (1964), 1-11.

24. Chipman, J. S. "On Least Squares With Insufficient Observations," American Statistical Association Journal, December, 1964.

25. Chipman, J. S., and Rao, M. M. "Projections, Generalized Inverses, and Quadratic Forms," Journal of Mathematical Analysis and Applications, 9, 1-11 (1964).

26. Cline, R. E. "On the Computation of the Generalized Inverse A, of an Arbitrary Complex Matrix A," Preliminary Report. Lafayette, Illinois: Purdue University, Statistical and Computing Laboratory, 1958.

27. Cline, R. E. "Note on the Generalized Inverse of the Product of Matrices," SIAM Review, 6 (January, 1964), 57-58.

28. Cline, R. E. "Representations for the Generalized Inverse of Matrices with Applications in Linear Programming." Doctoral dissertation, Purdue University, 1963. See also Journal of SIAM, 12 (September, 1964), 55-59.

29. Cline, R.E. "Representations for the Generalized Inverse of Sums of Matrices," Journal of SIAM Numerical Analysis, Series B, 2 (1965), 99-114.

30. Cline, R. E. "Representations for the Generalized Inverse of a Partitioned Matrix," Journal of Society of Industrial Applied Mathematics, 12 (1964), 588-600.

31. Decell, H. P., Jr. "An Application of Generalized Matrix Inversion to Sequential Least Squares Parameter Estimation" NASA TN D-2830.

32. Decell, H. P., Jr. "An Alternate Form for the Generalized Inverse of an Arbitrary Complex Matrix," SIAM Review, 7 (July, 1965), 356-358. (NASA TN D-2748).

33. Decell, H. P., Jr. "An Application of the Cayley-Hamilton Theorem to Generalized Matrix Inversion," SIAM Review, 7 (October, 1965), 526-528.

34. den Broeder, G. G., and Charnes, A. "Contributions to the Theory of a Generalized Inverse for Matrices," ONR Research Memo. No. 39. Evanston, Illinois: Northwestern University, The Technological Institute, 1962.

35. Dent, B. A., and Newhouse, J. "Polynomials Orthogonal over a Discrete Domain," SIAM Review, 1 (1959), 55-59.

36. Desoer, C. A., and Whalen, B. H. "A Note on Pseudo-Inverses," SIAM Journal, 11 (June, 1963), 442.

37. Drazin, M. P. "Pseudo-Inverses in Associative Rings and Semi-Groups," American Mathematical Monthly, 65 (September, 1958), 506-514.

38. Dwyer, P. S., and Macphail, M. S. "Symbolic Matrix Derivatives," Annals of Mathematical Statistics, Vol. 19, 517-534, (1948).

39. Faddeev, D. K., and Faddeeva, V. N. Computational Methods of Linear Algebra. San Francisco, California: W. H. Freeman Company, 1963, 260-265.

40. Florentin, J. J. "Optimal Control of Continuous Time, Markov, Stochastic Systems," Journal of Electronics Control, 10 (1961), 473-488.

41. Frame, J. S. "Matrix Operations and Generalized Inverses," IEEE Spectrum (March, 1964), 209-220.

42. Friedman, B. "Principles and Techniques of Applied Mathematics," Wiley, 1956.

43. Gainer, P. A. "A Method for Computing the Effect of an Additional Observation on a Previous Least-Squares Estimate," NASA TN D-1599.

44. Gantmacher, F. R., (Translated by J. L. Brennen), Applications of the Theory of Matrices, Interscience Publishers, Inc., New York, 1959.

45. Gateley, W. Y. "Application of the Generalized Inverse Concept to the Theory of Linear Statistical Models," Doctoral Dissertation, University of Oklahoma, 1962.

46. Geldman, A. J., and Zelen, M. "Weak Generalized Inverses and Minimum Variance Linear Unbiased Estimation," Journal of Research (National Bureau of Standards), 68-B (October-December, 1964).

47. Golub, G., and Kahan, W. "Calculating the Singular Values and Pseudo-Inverse of a Matrix," Journal of SIAM Numerical Analysis, Series B, 2 (1965), 205-224.

48. Graybill, F. A. "An Introduction to Linear Statistical Models" Vol. 1, McGraw Hill Book Co., 1961.

49. Greville, T. N. E. "The Pseudo-Inverse of a Rectangular Matrix and Its Applications to the Solution of Systems of Linear Equations," SIAM Review, 1 (January, 1959), 38-43.

50. Greville, T. N. E. "Some Applicati-ns of the Pseudo-Inverse of a Matrix," SIAM Review, 2 (January, 1960), 15-22.

51. Hestenes, M. R. "Inversion of Matrices by Biorthogonalization and Related Results," Journal of SIAM, 6 (1958), 84.

52. Hestenes, M. R. "Relative Hermitian Matrices," Pacific Journal of Mathematics, 1 (1961), 225-245.

53. Householder, A. The Theory of Matrices for Numerical Analysis. New York: Blaisdell Publishing Co., 1964.

54. Ijiri, Yuji "On the Generalized Inverse of an Incidence Matrix," Journal of Society of Industrial Applied Mathematics, Vol. 13, No. 3, September, 1965.

55. Kalman, R. E. "Contributions to the Theory of Optimal Control," Bol. Soc. Mat. Mexicana (1960), 102-119.

56. Kalman, R. E. "New Methods and Results in Linear Prediction and Filtering Theory," Technical Report No. 61-1. Baltimore, Maryland: RIAS, 1961.

57. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering, March, 1960, 35-44.

58. Koop, J. C. "Generalized Inverse of a Singular Matrix," _Nature_, 198 (June 8, 1963), 1019-1020.

59. Koop, J. C. "Generalized Inverse of a Singular Matrix," _Nature_, 200 (November 16, 1963), 716.

60. Lanczos, C. "Linear Systems in Self-Adjoint Form," _American Mathematical Monthly_, 65 (1958), 665-679.

61. Lass, H., and Solloway, C. B. "A Note on the Secular Equation of the Product of Two Matrices," _American Mathematical Monthly_, Vol. 68 #9, November, 1961, 906-907.

62. Lewis, T. O. "Application of the Theory of Generalized Matrix Inversion to Statistics," Doctoral Dissertation, The University of Texas (1966).

63. Magness, T. A. and McGuire, T. B. "Comparison of Least Squares and Minimum Variance Estimates on Regression Parameters," Vol. 33, No. 2, June, 1962, _The Annal of Mathematical Statistics_.

64. Marsaglia, G. "Conditional Means and Variances of Normal Variables with Singular Covariance Matrix," _American Statistical Journal_ (December, 1964), 1203-1204.

65. Moore, E. H. "On the Reciprocal of the General Algebraic Matrix," (Abstract), _Bulletin of the American Mathematical Society_, 26 (1960), 394-395.

66. Moore, E. H. "General Analysis, Part I," _Memoirs of the American Philosophical Society_, 1 (1935).

67. Munn, W. D. and Penrose, R. "A Note on Inverse Semi-Groups," Proceedings of the Cambridge Philosophical Society (England), 51 (1955), 396-399.

68. Odell, P. L. and Scroggs, J. E. "An Alternate Definition for the Pseudo-Inverse of a Matrix," to appear in SIAM Journal (1966).

69. Paige, L. J. and Swift, J. D. Elements of Linear Algebra. Boston: Ginn and Co., 1961.

70. Penrose, R. "A Generalized Inverse for Matrices," Proceedings of the Cambridge Philosophical Society (England), 51 (1955), 406-413.

71. Penrose, R. "On Best Approximate Solutions of Linear Matrix Equations," Proceedings of the Cambridge Philosophical Society (England), 52 (1956), 17-19.

72. Perlis, S. Theory of Matrices. Reading, Massachusetts: Addison-Wesley Publishing Co., Inc., 1958.

73. Price, C. M. "The Matrix Pseudo-Inverse and Minimal Variance Estimates," SIAM Review, Vol. 6, No. 2, April, 1964.

74. Pyle, L. D. "Generalized Inverse Computations Using the Gradient Projection Method," Journal of the Association for Computing Machinery, 11 (1964), 422-428.

75. Rado, R. "Note on Generalized Inverses of Matrices," Proceedings of the Cambridge Philosophical Society (England), 52 (1956), 600-601.

76. Rainbolt, M. B. "Sequential Least Squares Parameter Estimation" (Unpublished Report)

77. Rao, C. R. "A Note on a Generalized Inverse of a Matrix with Applications to Problems of Mathematical Statistics," Journal of the Royal Statistical Society (England), 24 (1962), 152-158.

78. Rosen, J. B. "The Gradient Projection Method for Nonlinear Programming, Part I: Linear Constraints," Journal of SIAM, 8 (1960), 181-217.

79. Rosen, J. B. "The Gradient Projection Method for Nonlinear Programming, Part II: Nonlinear Constraints," Journal of SIAM, 9 (1961), 514-532.

80. Rosen, J. B. "Minimum and Basic Solutions to Singular Linear Systems," Journal of SIAM, 12 (1964), 156-162.

81. Schulz, G. "Iterative Berechnung der Reziproken Matrix," Z. Angen. Math. Mech., 13 (1933), 57-59.

82. Sheffield, R. D. "On Pseudo-Inverses of Linear Transformations in Banach Spaces," Oak Ridge National Laboratory Report, No. 2133, 1956.

83. Sheffield, R. D. "A General Theory of Linear Systems," American Mathematical Monthly, 65 (February, 1958), 109-111.

84. Speed, F. M. and Feiveson, A. "Testing Hypotheses and Generalized Inverse" (Unpublished Report).

85. Tapley, B. D. and Odell, P. L. "A Study of Optimum Method for Determining and Predicting Space Vehicle Trajectories and Control Programs," Texas Center for Research Memorandum, No. 1, April 15, 1965, 43-62.

86. Tseng, Y. "Generalized Inverses of Unbounded Operators between Two Unitary Spaces," Dokl. Akad. Nauk. SSSR (N.S.), 67 (1949), 431-434. (Reviewed in Mathematical Review, 11 (1950), 115.)

87. Tseng, Y. "Properties and Classifications of Generalized Inverses of Closed Operators," Dokl. Akad. Nauk. SSSR (N.S.), 67 (1949), 607-610. (Reviewed in Mathematical Review, 11 (1950), 115.)

88. Tseng, Y. "Virtual Solutions and General Inversions," Uspehi. Mat. Nauk. (N.S.), 11 (1956), 213-215. (Reviewed in Mathematical Review, 18 (1957), 749.)

89. von Neumann, J. "On Regular Rings," Proceedings of the National Academy of Science, USA, 22 (1936), 707-713.

90. Zelen, M. "The Role of Constraints in the Theory of Least Squares," MRC Technical Summary Report No. 314. Madison, Wisconsin: University of Wisconsin, Mathematics Research Center, United States Army, 1962.

91. Zimmule, Don "Techniques for Computing the Pseudo-Inverse of a Matrix" (Unpublished Manuscript).